

Sistemi Intelligenti Algoritmi genetici

Alberto Borghese

Università degli Studi di Milano
Laboratory of Applied Intelligent Systems (AIS-Lab)
Dipartimento di Scienze dell'Informazione
borgnese@dsi.unimi.it



A.A. 2006-2007

1/55

<http://homes.dsi.unimi.it/~borgnese/>



Sommario



- **Algoritmi genetici**
- Ottimizzazione evolutivistica
- Applicazione delle strategie evolutive alla visione

A.A. 2006-2007

2/55

<http://homes.dsi.unimi.it/~borgnese/>



Evoluzione



- Processo discontinuo su un arco temporale ampio e su un numero di individui grande, che porta la specie al successo.
- Computazione evoluzionistica è un termine generico che indica una gamma di sistemi di risoluzione di problemi di ottimizzazione che emulano l'evoluzione naturale nel ricercare la soluzione migliore.
- Algoritmi genetici, Ottimizzazione evolutiva. Programmazione Genetica.

A.A. 2006-2007

3/55

<http://homes.dsi.unimi.it/~borghese/>

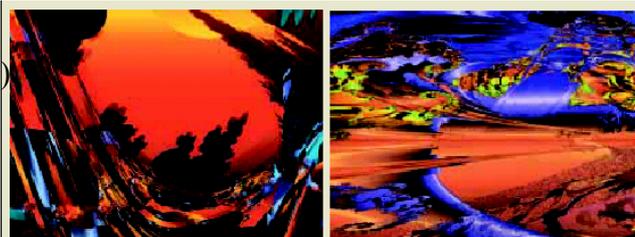


Alcune pietre miliari



- 1960: Ingo Rechenberg introduce l'idea della computazione evoluzionistica nel suo lavoro "Evolution strategies"
- L. Fogel 1962 (San Diego, CA): Programmazione evoluzionistica
- I.Rechenberg, H.P.Schwefel, 1972, Ottimizzazione evolutiva.
- 1975: John Holland inventa gli algoritmi genetici e pubblica il libro "Adaption in Natural and Artificial Systems"
- 1992: John Koza propone gli algoritmi genetici per fare evolvere programmi software che effettuano specifici compiti. Tale metodo viene chiamato da lui Programmazione Genetica

- Applicazioni varie (Steven Rooke, 1991)



"In the Beginning" (left); "The High Plains of Kilimanjaro" (right).

A.A. 2006-2007



Algoritmi Genetici



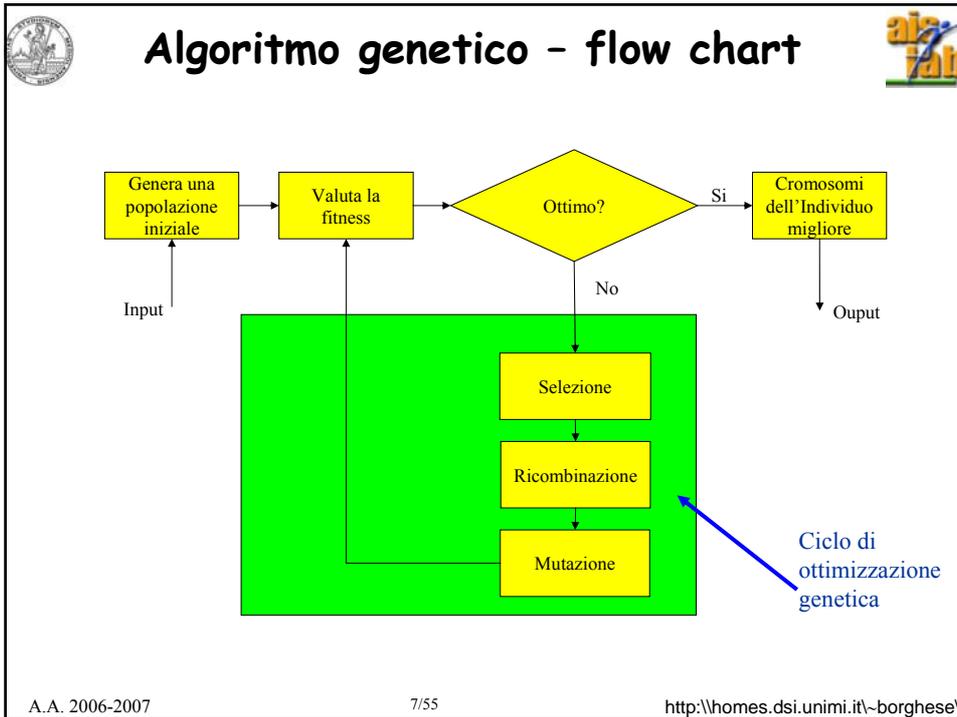
- **Cromosoma:** La codifica di una possibile **soluzione** per un dato problema di solito rappresentata tramite un array di bit o caratteri. Può essere associato ad una variabile, più variabili possono essere associate a diversi cromosomi.
- Corpuscolo all'interno della cellula contenente un tratto di DNA. Il DNA è una struttura allungata, a doppia elica, costituita da filamenti di basi affacciate e connesse a coppie: Adenina-Timina, Citosina-Guanina o Uracile.
- **Gene.** Un singolo bit o un insieme di bit, parte della soluzione (e.g. parte di una variabile).
- Sequenza di basi del DNA che codifica un'informazione (e.g. colore occhi), sottoparti di un cromosoma.
- **Ricombinazione.** Vengono generate nuove soluzioni mescolando due soluzioni (cromosomi).
- Nel generare il cromosoma di un figlio, tratti omologhi del cromosoma dei genitori si scambiano materiale genetico.
- **Mutazione.** Cambiamenti casuali nella soluzione.
- "Errore" che avviene durante la replicazione del DNA generato dalla ricombinazione.



Ottimizzazione genetica



- Codifica della potenziale soluzione di uno specifico problema in una struttura dati associabile ad un cromosoma artificiale.
- Applicazione iterativa del processo di ricombinazione e mutazione a queste strutture.
- Valutazione del risultato mediante funzione di **fitness**.
- **Fitness:** misura del successo di un individuo della specie.



- ## Algoritmo genetico - i passi
1. [Inizializzazione] Genera una popolazione random di n cromosomi (soluzioni plausibili - genitori)
 2. [Fitness] Valuta la fitness $f(x)$ di ciascun cromosoma x nella popolazione corrente (bontà della soluzione associata al cromosoma).
 3. [Nuova popolazione] Crea una nuova popolazione (prole) attraverso:
 - (a) [Selezione] Seleziona i cromosomi genitori in base alla loro fitness (criterio)
 - (b) [Ricombinazione] Con una certa probabilità di ricombinazione incrocia due genitori per generare un nuovo figlio.
 - (c) [Mutazione] Con una certa probabilità di mutazione cambia i geni che costituiscono il figlio
 - (d) [Accettazione] Introduce il nuovo figlio nella popolazione.
- La soluzione è ricavata dall'elemento con la miglior fitness.
- A.A. 2006-2007 8/55 http:\\homes.dsi.unimi.it\\~borghese\



Gli elementi da definire



- Come creare i cromosomi e che tipo di codifica utilizzare
- Come selezionare i genitori per la ricombinazione nella speranza che i genitori migliori produrranno migliore prole
- Come definire la ricombinazione e la mutazione



Codifica della soluzione



La soluzione è codificata nel cromosoma.

- Un cromosoma deve contenere l'informazione sulla soluzione che rappresenta
- La codifica dipende principalmente dal problema da risolvere (e.g. numeri binari, numeri reali, permutazione, alberi di parsing,...)



Codifica binaria



- La codifica binaria è la più comune codifica implicita (principalmente perché la prima ricerca di GA aveva utilizzato tale codifica)
 - Codifica binaria: ogni cromosoma è una stringa di bits (1 o 0)
 - L'implementazione di operatori genetici risulta essere immediata
 - Tuttavia non è sempre naturale per molti problemi

Esempio: problema dello zaino

Date un certo numero di oggetti caratterizzati ciascuno da un valore e una dimensione, massimizzare il valore degli oggetti in uno zaino che ha una capacità fissata

In questo caso, ogni bit può rappresentare il fatto che un'oggetto è inserito o no nello zaino.

Cromosoma 1

1101100100110110

Cromosoma 2

1110111000011110

A.A. 2006-2007

11/55

<http://homes.dsi.unimi.it/~borghese/>



Codifica binaria (alternativamente)



- La codifica binaria è la più comune codifica implicita (principalmente perché la prima ricerca di GA aveva utilizzato tale codifica)
 - Codifica binaria: ogni cromosoma è una stringa di bits (1 o 0)
 - L'implementazione di operatori genetici risulta essere immediata
 - Tuttavia non è sempre naturale per molti problemi

Esempio: ottimizzazione,

La soluzione viene codificata in un certo numero di bit. Variando i bit più significativi varia il range della soluzione, con i bit meno significativi, si sposta di poco la soluzione.

- Elitismo.** Quando si selezionano i nuovi padri si tiene conto della loro fitness, implicitamente si salvano via via le soluzioni migliori. Nel caso della codifica binaria si può introdurre un "blocco" o una "resistenza" alla modifica dei geni migliori.

A.A. 2006-2007

12/55

<http://homes.dsi.unimi.it/~borghese/>



Codifica esplicita



La codifica esplicita o diretta è indicata in problemi dove le entità coinvolte sono caratterizzate da una certa complessità

- Il cromosoma è una sequenza di valori connessi al problema (numeri, caratteri, stringhe di caratteri, ...)
- Rappresenta la scelta migliore per ottimizzazione in spazi multidimensionali e/o multimodali. Gli operatori di ricombinazione e mutazione richiedono specifici controlli di consistenza

Esempio: addestramento di un agente (determinazione della policy)

A	B	D	H	Y	V	S	V
---	---	---	---	---	---	---	---

I numeri reali nel cromosoma rappresentano le probabilità di scegliere le azioni in un certo stato.

2.5678	1.4361	3.3426	7.8761
--------	--------	--------	--------

open	walk	back	close
------	------	------	-------



Codifica di permutazioni



La codifica di permutazione è utilizzata per problemi di ordinamento

- Il cromosoma è una stringa di numeri che rappresenta la posizione in una sequenza
- La ricombinazione e la mutazione devono essere determinate in modo da produrre prole consistente

Esempio: problema del commesso viaggiatore (Traveling salesman problem)

- È dato un insieme di città e le corrispondenti distanze a coppie. Il commesso viaggiatore deve visitarle tutte ma non vuole viaggiare più del necessario. Scopo: Trovare una sequenza di città che minimizza la distanza percorsa.
 - ◆ Il cromosoma descrive l'ordine delle città visitate

Cromosoma 1

1	5	7	8	3	5	1	3	1	0	1	1	6	2	1	1	4	2	4	6	9
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Cromosoma 2

2	9	1	4	1	1	5	8	1	5	1	3	6	1	2	1	6	7	3	1	0	4
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---



Selezione



- In accordo alla teoria evolutiva di Darwin il migliore cromosoma sopravvive per creare nuova prole. Per selezionare il migliore cromosoma esistono diversi criteri:
 - ◆ Ordinamento (priorità diretta ai cromosomi che si sono espressi meglio)
 - ◆ Casualità (metodo della roulette, solo tra i cromosomi migliori)
 - ◆ Probabilità crescente con l'incremento di espressione
 - ◆ ...

In particolare è possibile dare una probabilità di selezione proporzionale alla fitness (Roulette)



Selezione tramite roulette



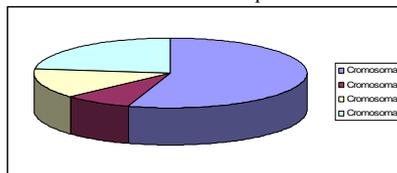
Gli individui sono selezionati proporzionalmente alla loro fitness. Migliore essa è e più alta è la probabilità di selezione

1. Si immagini una roulette dove sono sistemati tutti i cromosomi della popolazione
2. La dimensione della sezione nella roulette è proporzionale al valore di fitness di ciascun cromosoma
3. La pallina viene lanciata all'interno della roulette e il cromosoma in corrispondenza del quale si ferma è quello selezionato

Esempio:

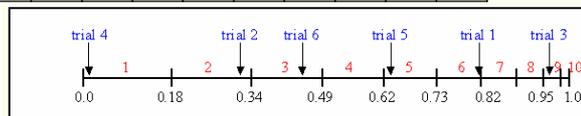
Generazione casuale di 6 numeri (0 1)

0.81, 0.32, 0.96, 0.01, 0.65, 0.42



Numero di individui	1	2	3	4	5	6	7	8	9	10	11
Fitness	3.0	2.7	2.4	2.1	1.8	1.5	1.2	0.9	0.6	0.3	0.0
Probabilità di selezione	0.18	0.16	0.15	0.13	0.11	0.09	0.07	0.06	0.03	0.02	0.00

Cromosoma 1:
 $P(\text{figlio} = 1) = 3.0 / 16.5 = 0.1818$



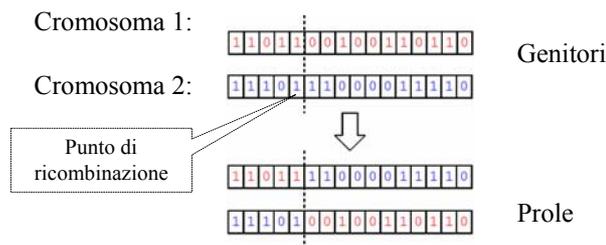


Ricombinazione (binaria)



Il processo di ricombinazione di base opera su geni selezionati dal cromosoma genitore e crea nuova prole:

1. Selezione random di un **punto di ricombinazione** all'interno del cromosoma
2. Copia tutti i geni precedenti questo punto dal primo genitore e poi copia tutti i geni successivi a questo punto dal cromosoma del secondo genitore



A.A. 2006-2007

17/55

<http://homes.dsi.unimi.it/~borghese/>



Estensione della ricombinazione binaria



Per la codifica binaria esistono diverse modalità:

- **Ricombinazione da un singolo punto:** selezionato un punto all'interno del cromosoma, la stringa binaria compresa tra l'inizio e il punto di ricombinazione è copiata dal primo genitore, il resto è copiata dal secondo genitore
- **Ricombinazione da due punti:** selezionati i due punti all'interno del cromosoma, la stringa binaria compresa tra l'inizio e il primo punto di ricombinazione è copiata dal primo genitore, la parte compresa tra il primo e il secondo punto è copiata dal secondo genitore, la parte compresa tra il secondo punto e la fine del cromosoma è copiata ancora dal primo genitore
- **Ricombinazione uniforme:** i bit sono copiati casualmente dal primo o dal secondo genitore
- **Ricombinazione aritmetica:** operatori algebrici (e.g., logica AND, OR, ..). Altre funzioni booleane possono essere utilizzate.

A.A. 2006-2007

18/55

<http://homes.dsi.unimi.it/~borghese/>



Mutazione (binaria)



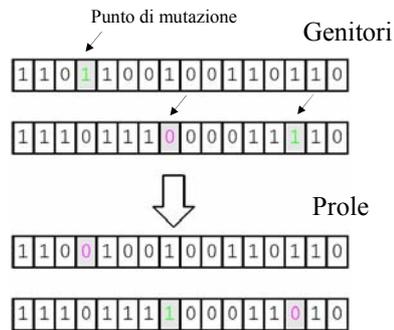
Dopo la ricombinazione interviene il processo di mutazione

Nella mutazione di base cambia in modo casuale un numero contenuto di bits da 0 a 1 o viceversa da 1 a 0

Problemi:

Come preservare i cromosomi che danno soluzioni già buone (Etilismo)?

Come conservare la consistenza della soluzione (e.g. problema del commesso viaggiatore)?



Sommario



- Algoritmi genetici
- **Ottimizzazione evolutiva**
- Applicazione delle strategie evolutive alla visione



I genitori della prima generazione



- Le soluzioni beneficiano di una codifica esplicita.
- I genitori iniziali sono scelti da una distribuzione statistica, solitamente uniforme
- La selezione dei genitori nei modelli più semplici di ES non è polarizzata solitamente – ciascun individuo ha la stessa probabilità di essere scelto
- Tra una generazione e la successiva avvengono gli eventi evolutivi:
 - ◆ Ricombinazione
 - ◆ Mutazione
 - ◆ Selezione



Ricombinazione in ES



Quali sono i genitori che diventano genitori nella generazione successiva?

Identificazione degli elementi per la produzione dei figli:

1. Estrae ad ogni iterazione (generazione) uno o più nuovi elementi.
2. Agisce sugli elementi correnti per determinare i nuovi genitori:
 - ◆ Combinando cromosomi corrispondenti agli elementi (nel caso comune di più elementi)
 - ☞ **Ricombinazione discreta**
Individuo 1: [12 25 5] estrazione 1: 2 2 1 Figlio 1: [123 4 5]
Individuo 2: [123 4 34] estrazione 1: 1 2 1 Figlio 2: [12 4 5]
 - ☞ **Ricombinazione intermedia**
 $\text{Figlio} = \text{individuo1} + \alpha(\text{individuo2} - \text{individuo1})$
 - ◆ Selezionando uno tra gli elementi disponibili.
 - ☞ Esempio di soluzione a valori discreti → elemento intermedio
 - ☞ Esempio di soluzione a valori reali: → posizione media.



Mutazione in ES

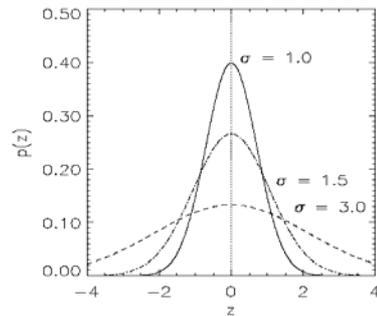


Genitore^{t+1} = Elemento^t + z

■ **Mutazione:** perturbare i valori aggiungendo un rumore casuale. Spesso è utilizzata una distribuzione Gaussiana a media nulla e deviazione standard σ , $N(0, \sigma) \rightarrow$
 $x_i^? = x_i + N(\sigma, 0)$

- Media nulla
- Deviazione standard σ

La deviazione standard σ è chiamata passo di mutazione ed è una proprietà comune a tutti i cromosomi.



Selezione in ES



μ - # genitori λ - # figli.

Metto in competizione padri e figli nel processo di selezione.

Strategia $(\mu + \lambda)$ – I genitori della generazione successiva vengono scelti tra i $\mu + \lambda$ elementi della popolazione attuale. E' una strategia elitista.

Strategia (μ, λ) – I genitori della generazione successiva vengono scelti solamente tra i λ figli. E' la strategia preferita, soprattutto con adattamento della matrice di covarianza. La strategia $(\mu + \lambda)$ tende a concentrare le soluzioni in luoghi che possono dare soluzioni sub-ottimali.

$\lambda \gg \mu$ e.g. $\lambda = 7 \mu$

I genitori selezionati vengono poi ricombinati per diventare gli elementi della generazione successiva.



Esempio I - (1+1) ES



- Obiettivo: minimizzare $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$
- Algoritmo: “two-membered ES” usando:
 - ◆ Vettori in \mathbb{R}^n direttamente come cromosoma
 - ◆ Dimensione della popolazione: 1
 - ◆ Non c'è ricombinazione.
 - ◆ Opera solo la mutazione che crea un singolo figlio
 - ◆ Valutazione della funzione per il padre e il figlio
 - ◆ Scelgo come padre della generazione successiva chi tra i due ha espresso una fitness migliore.

A cosa è equivalente?

Problema?



Variabili strategiche



Ogni gene di un cromosoma (variabile) ha associato un'ampiezza di mutazione personalizzata, mutazione scorrelata.

La **mutazione** dipende da σ :

Figlio^t = Elemento^t + z, con $z = N(0, \sigma)$ → $\mathbf{x}^{(t)'}(n) = \mathbf{x}^{(t)}(n) + \mathbf{z}$
quanto vale σ ?

Cromosomi: $[x_1, \dots, x_n; w]$

$\{x_1, x_n\}$ rappresentano gli elementi della soluzione (ad esempio gli elementi di un vettore), $\{w\}$ rappresentano le **variabili strategiche**, ad esempio σ , che determinano l'ampiezza della mutazione.

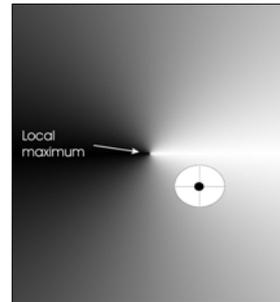


Spazio isotropo di mutazione



La mutazione isotropa genera figli distribuiti uniformemente all'interno di un'ipersfera nello spazio delle soluzioni

$$\mathbf{x}'(n) = \mathbf{x}(n) + N(0, \sigma)$$



A.A. 2006-2007

27/55

<http://homes.dsi.unimi.it/~borghese/>



Adattamento dell'ampiezza della regione di mutazione - I



$$\mathbf{x}'(n) = \mathbf{x}(n) + \sigma^2(n) \cdot N(0,1)$$

- σ viene aggiornata run-time con la regola euristica "1/5 success rule".

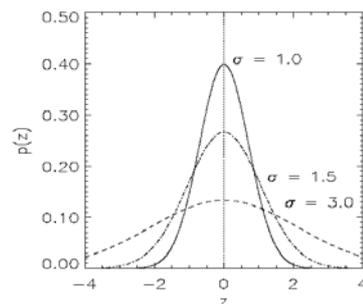
Questa regola ricalcola σ ogni k epoche tramite:

$$\rightarrow \sigma = \sigma / c \quad \text{se } p_s > 1/5$$

$$\rightarrow \sigma = \sigma \cdot c \quad \text{se } p_s < 1/5$$

$$\rightarrow \sigma = \sigma \quad \text{se } p_s = 1/5$$

dove p_s è la percentuale di mutazioni con successo, $0.8 \leq c \leq 1$.



$$\sigma^2(n) = \sigma(n) \cdot k(n) = \mathbf{I}(\sigma)$$

$$\mathbf{x}'(n) = \mathbf{x}(n) + \mathbf{I}(\sigma) N(0,1) = \mathbf{x}(n) + N(0, \sigma)$$

A.A. 2006-2007

28/55

<http://homes.dsi.unimi.it/~borghese/>



Esempio di mutazione per un singolo gene



Valuto se il figlio è migliore del padre. Supponiamo ES (1+10)

- Es: $k=10$
 - ◆ $k = 1$ -- fitness figlio > fitness padre
 - ◆ $k = 2$ -- fitness figlio > fitness padre
 - ◆ $k = 3$ -- fitness figlio < fitness padre
 - ◆ $k = 4$ -- fitness figlio > fitness padre
 - ◆ $k = 5$ -- fitness figlio < fitness padre
 - ◆ $k = 6$ -- fitness figlio > fitness padre
 - ◆ $k = 7$ -- fitness figlio < fitness padre
 - ◆ $k = 8$ -- fitness figlio < fitness padre
 - ◆ $k = 9$ -- fitness figlio > fitness padre
 - ◆ $k = 10$ -- fitness figlio > fitness padre

- $4/10 > 1/5$ quindi $\sigma = \sigma / c$.

σ decresce esponenzialmente ad ogni generazione. In aggiunta, se $p_s \ll 1/5$, si somma uno scalino crescente o decrescente.

A.A. 2006-2007

29/55

<http://homes.dsi.unimi.it/~borghese/>



Adattamento dell'ampiezza della regione di mutazione - II



- Mutazione correlata con un valore di σ uguale per ogni cromosoma.
 - ◆ Cromosomi: $[x_1, \dots, x_n, \sigma]$
 - ◆ $x'_i(n) = x_i(n) + \sigma'(n) \cdot N_i(0,1)$
 - Voglio determinare una legge per cui l'ampiezza della regione di ricerca diminuisca con il procedere dell'evoluzione. Una possibilità è quello di restringere esponenzialmente l'ampiezza della regione mediante:
 - ◆ $\sigma'(n) = \sigma(n) \cdot e^{[-\tau'N(0,1)]}$
 - 1 parametro aggiuntivo.
 - ☞ τ' learning rate globale
 - ◆ Valori tipici: $\tau' \propto [1/(2n^{1/2})]^{1/2}$ $\tau' > 0$,
dove n è il numero di generazioni.
- σ decresce esponenzialmente con il numero di generazioni.
- ◆ Limiti: $\sigma'_i > \epsilon_0$

A.A. 2006-2007

30/55

<http://homes.dsi.unimi.it/~borghese/>



Mutazione - σ personalizzato



- Inserisco valori diversi di σ per diverse variabili, cioè per i diversi cromosomi. Concedo che lo spazio delle soluzioni sia esplorato maggiormente lungo alcune direzioni rispetto altre.
- Mutazione scorrelata con un valore di σ_i associato ad ogni cromosoma.
 - ◆ Cromosomi: $[x_1, \dots, x_n, \sigma_1, \dots, \sigma_n]$
 - ◆ $x'_i(n) = x_i(n) + \sigma'_i(n) \cdot N_i(0,1)$

Come aggiornare in modo diverso l'ampiezza della mutazione associata ai diversi cromosomi?



Mutazione - σ personalizzato

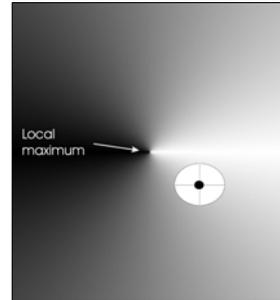


- Inserisco valori diversi di σ per diverse variabili, cioè per i diversi cromosomi. Concedo che lo spazio delle soluzioni sia esplorato maggiormente lungo alcune direzioni rispetto altre.
- Mutazione scorrelata con un valore di σ_i associato ad ogni cromosoma.
 - ◆ Cromosomi: $[x_1, \dots, x_n, \sigma_1, \dots, \sigma_n]$
 - ◆ $\sigma'_i(n) = \sigma_i(n) \cdot e^{[-\tau' N(0,1) - \tau_i N_i(0,1)]}$, alternativa alla Rechenberg rule
 - ◆ $x'_i(n) = x_i(n) + \sigma'_i(n) \cdot N_i(0,1)$
- 2 parametri di learning:
 - ☞ τ' learning rate globale
 - ☞ τ_i learning rate relativo alla specifica variabile i-esima
- N - 1 variabili strategiche aggiuntive.
 - ◆ Valori tipici: $\tau_i \propto 1/(2n)^{1/2}$ e $\tau' \propto [1/(2n)^{1/2}]^{1/2}$ $\tau_i > 0$ $\tau' > 0$, dove n è il numero di generazioni.
 - ◆ Limiti: $\sigma'_i > \epsilon_0$



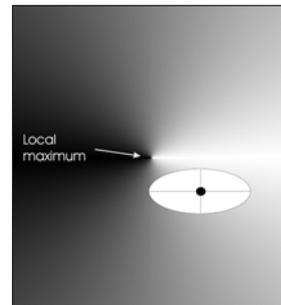
La mutazione isotropa genera figli distribuiti uniformemente all'interno di un'ipersfera nello spazio delle soluzioni

$$\mathbf{x}'(n) = \mathbf{x}(n) + N(0, \sigma)$$



La mutazione anisotropa genera figli distribuiti uniformemente all'interno di un'iperellissoide nello spazio delle soluzioni

$$\mathbf{x}'(n) = \mathbf{x}(n) + N(0, \mathbf{D})$$



Adattamento della mutazione avanzato



Favorisco le variazioni degli elementi della popolazione lungo le direzioni dove σ_i è maggiore.

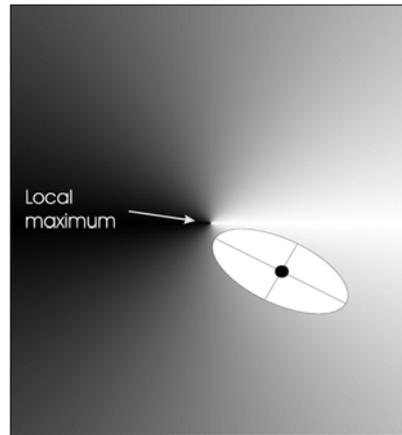
Vorrei trovare un modo di adattare l'ampiezza della regione di ricerca nelle diverse direzioni (variabili) in funzione della fitness (bontà) della soluzione.

Correlazione tra fitness e variabili.



Gli elementi, generati per mutazione sono statisticamente distribuiti all'interno dell'iperellissoide orientato lungo le direzioni principali della varianza.

Questo consente di adattare la regione di ricerca al meglio alla topologia della funzione obiettivo senza doverla determinare in modo esplicito.



σ correlato tra i vari cromosomi



- Mutazioni correlate; la correlazione viene rappresentata nella matrice di covarianza.
 - ◆ Cromosomi: $[x_1, \dots, x_n; \sigma_1, \dots, \sigma_n; \alpha_{12}, \alpha_{13}, \dots, \alpha_{N-1, N-1}]$.
 - ◆ La matrice di covarianza $C_{n \times n}$ gode di queste proprietà:
 - ☞ simmetrica
 - ☞ $c_{ii} = \sigma_i^2$
 - ☞ $c_{ij} = 0$ se i e j sono scorrelati.
 - ☞ $c_{ij} \neq 0, i \neq j, c_{ij}$ viene codificato dai parametri α_{ij}^2 .

$$C = B D^2 B'$$

B – matrice ortonormale

D – matrice diagonale



Ampiezza della Mutazione - σ correlato



Nel caso scorrelato, avevamo:

$$\sigma'(n) = \sigma(n) \cdot \mathbf{k}(n) = \mathbf{D}(n) \quad \mathbf{D}, \text{ matrice diagonale.}$$

$$\mathbf{x}'(n) = \mathbf{x}(n) + N(0, \mathbf{D})$$

■ Meccanismo di mutazione:

◆ $x'_i(n) = x_i(n) + \sigma'_i(n) \cdot N_i(0,1)$

◆ $\sigma'_i(n) = \sigma_i(n) \cdot e^{[-\tau' N(0,1) - \tau_i' N_i(0,1)]}$

◆ $\alpha'_{ij} = \alpha_{ij} + \beta \cdot N(0,1)$

◆ $\mathbf{x}' = \mathbf{x} + N(0, \mathbf{C}') = \mathbf{x} + N(0, \mathbf{B}\mathbf{D})$

☞ \mathbf{C}' è la matrice di covarianza \mathbf{C} dopo la mutazione dei valori α e σ .

◆ Valori tipici: $\tau_i \propto 1/(2n)^{1/2}$ e $\tau' \propto 1/(2n^{1/2})^{1/2}$ $\tau_i > 0$ $\tau' > 0$; and $\beta \approx 5$

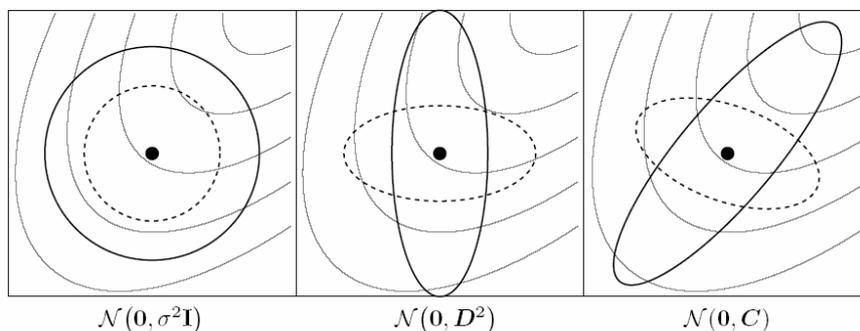
◆ Nota bene:

☞ Il valore di σ determina l'ampiezza della regione di ricerca nelle diverse direzioni (σ_i), matrice \mathbf{D} .

☞ La regione ellissoidale viene poi ruotata utilizzando la matrice di covarianza, matrice \mathbf{B} .



Sintesi





Il ciclo evolutivo per cromosomi correlati



Ricombinazione (metodi standard)
Mutazione
Selezione



Mutazione



$$\mathbf{x}_k^{(g+1)} = \mathbf{m}_j^{(g)} + \rho^{(g)} \mathbf{B} \mathbf{D} \mathbf{z}_k$$

$$\mathbf{z}_k = [z_1 = N(0,1), z_2 = N(0,1), \dots, z_n = N(0,1)]^T \quad \text{Gaussiana}$$

- Vengono generati λ figli, \mathbf{x}_k , a partire dalla mutazione di un genitore, \mathbf{m}_j .
- $\mathbf{x}^{(g+1)}$ sarà funzione dei padri selezionati
- $\rho^{(g)}$ è globale: global step size
- \mathbf{B} \mathbf{D} definiscono l'iper-ellissoide all'interno del quale effettuare le mutazioni.

Devo determinare: \mathbf{B} \mathbf{D} e ρ . Per questo utilizzo l' "evolution path"



Derandomizzazione tramite Evolution Path



Sia la matrice di covarianza C che il global step size ρ sono aggiornati indirettamente tramite il cosiddetto evolution path \mathbf{s} .

$$\mathbf{s}^{(g+1)} = (1 - \alpha_s) \mathbf{s}^{(g)} + \alpha_u \mathbf{B}^{(g)} \mathbf{D}^{(g)} \mathbf{z}_{i_{sel}}^{(g)} \quad \mathbf{z}_{i_{sel}} \text{ rappresenta il genitore migliore}$$

$$a_u^2 + (1 - a_s)^2 = 1$$

$$\text{Nuovo_valore} = a * \text{Vecchio_valore} + b * \text{Innovazione}$$

Cf. Value estimate in RL

s unisce tutti i genitori migliori. Dà un'idea dell'evoluzione. Path evolutivo all'interno dello spazio delle soluzioni. Connette i genitori migliori.



Aggiornamento della matrice di Covarianza



$$\text{Nuovo_valore} = a * \text{Vecchio_valore} + b * \text{Innovazione}$$

$$\mathbf{C}^{(g+1)} = (1 - \alpha_C) \mathbf{C}^{(g)} + \alpha_C \mathbf{S}^{(g+1)}$$

$$\mathbf{S}^{(g+1)} = \mathbf{s} \mathbf{s}^T \quad \text{Matrice di covarianza calcolata sul path evolutivo}$$

Cosa vuol rappresentare questa equazione di aggiornamento?

Tanto più la covarianza rimane stabile, tanto più \mathbf{C} cresce \rightarrow \mathbf{D} cresce e \mathbf{B} rimane stabile \rightarrow \mathbf{x}_k si muove più velocemente. $\mathbf{x}_k^{(g+1)} = \mathbf{m}_j^{(g)} + \rho^{(g)} \mathbf{B} \mathbf{D} \mathbf{z}_k$

Ha effetto anche su ρ



Adattamento del global step size ρ



$\sigma_\rho^{(g+1)}$ Evolution path normalizzato (contiene solamente informazioni sulla direzione).

$$\mathbf{s}_p^{(g+1)} = (1 - \alpha_s) \mathbf{s}_p^{(g)} + \alpha_s \mathbf{B}^{(g)} \mathbf{z}_{i_{sel}}^{(g)}$$

$$\bar{e}_n = \left(1 - \frac{1}{4n} + \frac{1}{21n^2} \right) \quad \bullet \quad \text{Approssimazione al 2° ordine del valore atteso della distribuzione della lunghezza dei vettori estratti da } M(\mathbf{0}, \mathbf{I})$$

$$\rho^{(g+1)} = \rho^{(g)} e^{\gamma \left(\left\| \mathbf{s}_p^{(g+1)} \right\| - \bar{e} \right)} \quad \bullet \quad \rho \text{ decresce quando la direzione dell'evolution path cambia frequentemente. Quando una stessa direzione è ripetutamente selezionata } \rho \text{ è incrementato.}$$

$$\gamma = 0.1$$

Se l'evoluzione cambia spesso direzione, meglio generare figli più vicini (e saranno sparsi lungo tutte le direzioni).

A.A. 2006-2007

43/55

<http://homes.dsi.unimi.it/~borghese/>



I parametri secondari: riassumendo



$$\mathbf{x}_k^{(g+1)} = \mathbf{m}^{(g)} + \rho^{(g)} \mathbf{B} \mathbf{D} \mathbf{z}_k$$

$$\mathbf{s}^{(g+1)} = (1 - a_s) \mathbf{s}^{(g)} + a_u \mathbf{B}^{(g)} \mathbf{G}^{(g)} \mathbf{z}_{i_{sel}}^{(g)}$$

$$\mathbf{s}_p^{(g+1)} = (1 - \alpha_s) \mathbf{s}_p^{(g)} + \alpha_s \mathbf{B}^{(g)} \mathbf{z}_{i_{sel}}^{(g)}$$

$$a_u^2 + (1 - a_s)^2 = 1$$

$$\mathbf{C}^{(g+1)} = (1 - \alpha_C) \mathbf{C}^{(g)} + \alpha_C \mathbf{S}^{(g+1)}$$

$$\mathbf{S}^{(g+1)} = \mathbf{s} \mathbf{s}^T$$

$$\rho^{(g+1)} = \rho^{(g)} e^{\gamma \left(\left\| \mathbf{s}_p^{(g+1)} \right\| - \bar{e} \right)}$$

Rimangono da determinare: α_C , α_S , α_u

A.A. 2006-2007

44/55

<http://homes.dsi.unimi.it/~borghese/>



Aggiornamento di α_s

Regola l'aggiornamento di ρ . Questo deve essere particolarmente sensibile all'innovazione, in modo da pilotare le generazioni future dove più serve.

$$\alpha_s = 1/\sqrt{n}$$

Dove n è la dimensionalità dello spazio della popolazione

$$\alpha_C = 1/n^2$$

Il numero di parametri liberi della matrice di covarianza è dell'ordine di $n^2 \rightarrow 1/n^2$.

Rappresentano due fenomeni diversi: r è sensibile alle variazioni rapide della fitness della popolazione, mentre C è più sensibile alle variazioni lente, filtrando perciò le variazioni brusche occasionali.

- Es: $n=10$
 - ◆ $\alpha_s=0.01, \alpha_u=0.72$
- La matrice di covarianza ha un *time span* molto più ampio dell'evolution path $O(n^2)$
- L'evolution path può subire cambiamenti veloci mentre la matrice di covarianza ha bisogno di molto più tempo per fissare la memoria della direzione (deve essere selezionata molte volte)

A.A. 2006-2007

45/55

<http://homes.dsi.unimi.it/~borghese/>



selezione all'epoca g

Calcola evolution path normalizzato
 s_p

Aggiorna evolution path normalizzato

Aggiorna global step size ρ

Aggiorna evolution path s

Calcola matrice di covarianza corrente
 $S=ss^T$

Aggiorna matrice di covarianza C

Schema di flusso dell'adattamento della mutazione

A.A. 2006-2007

46/55

<http://homes.dsi.unimi.it/~borghese/>



Sommario



- Algoritmi genetici
- Ottimizzazione evolutiva
- **Applicazione delle strategie evolutive alla visione**



Problema 1: Calibrazione geometrica di una coppia stereo di telecamere



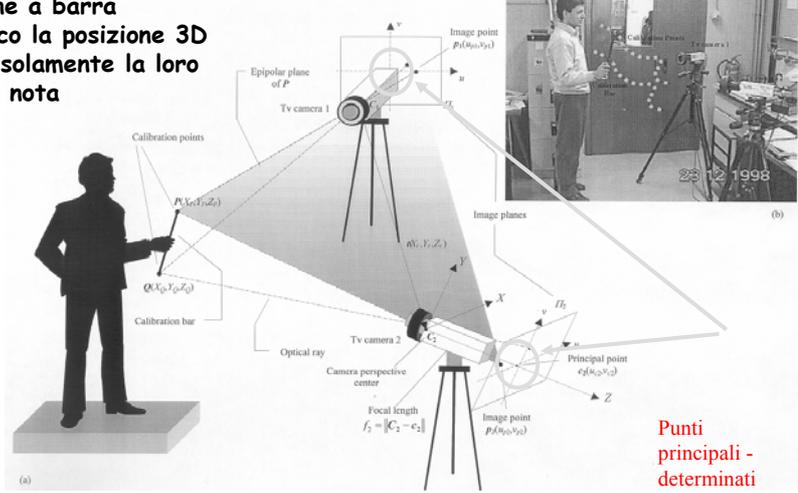
- Determinazione dei parametri geometrici del set-up:
 - ◆ Posizione e orientamento relativi di una coppia di camera.
 - ◆ Lunghezza focale e punto principale di ciascuna camera.
- Utilizzando la geometria epipolare
 - ◆ Relazione tra punti omologhi sulle due telecamere
 - ◆ Matrice fondamentale 3x3 (7 parametri liberi)
 - ≈ 3 rotazioni, 3 traslazioni a meno di una fattore di scala (2Dofs), 2 lunghezze focali
 - ◆ **Consente di calcolare 7 parametri in forma chiusa utilizzando la geometria proiettiva.**
- Punti principali
 - Stimo a posteriori i 2 punti principali, c_1 e c_2 , tramite **Ricerca evolutiva**

Spazio della ricerca quadridimensionale, $c_1[x_1, y_1]$ e $c_2[x_2, y_2]$ ("mondo" della popolazione).

Il set-up




Calibrazione a barra
 Non conosco la posizione 3D dei punti, solamente la loro distanza è nota



Punti principali - determinati mediante ES

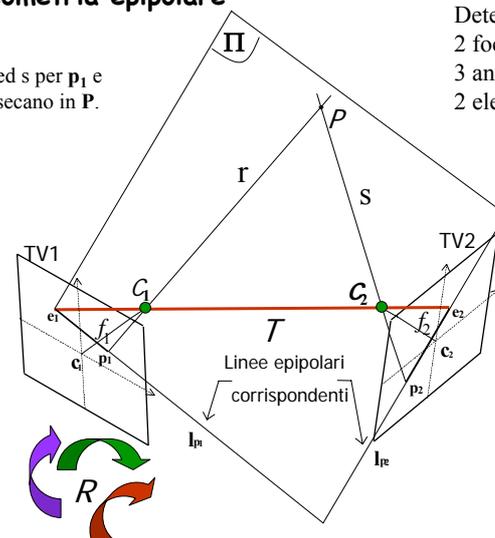
A.A. 2006-2007 49/55 http://homes.dsi.unimi.it/~borghese/

Geometria epolare




Le rette r ed s per p_1 e p_2 si intersecano in P .

Determino 7 parametri:
 2 focali
 3 angoli di orientamento
 2 elementi del vettore posizione relativa



I parametri sono contenuti nella matrice F

$$\overline{PC_2} \cdot \overline{C_1 C_2} \wedge \overline{PC_1} = p_2^T K_2^{-T} R(T \wedge) K_1^{-1} p_1 = 0 \Rightarrow p_2^T F p_1 = 0$$

p_1 e p_2 sono misurati rispetto a c_1 e c_2 non noti.

A.A. 2006-2007 50/55 http://homes.dsi.unimi.it/~borghese/



Soluzione tramite "Evolution strategies"



Procedura: ES(1,10)

1. scelgo una possibile soluzione per la coppia di punti principali
2. calcolo 7 parametri geometrici in forma chiusa, tramite matrice fondamentale.
3. effettuo la ricostruzione 3D e calcolo il fattore di scala (1 parametro).
4. Valuto la fitness della soluzione.

Funzione fitness:
$$J = \alpha \sqrt{\frac{1}{N} \sum_{j=1}^N (d_j - L)^2} + \beta \sqrt{\frac{1}{N} \sum_{j=1}^N \left(\frac{e_{\text{int } j_r}^2 + e_{\text{int } j_o}^2}{2} \right)}$$

Errore sulla lunghezza della barra. \rightarrow \leftarrow Errore di intersezione (tra r ed s)

$$\min_{S_2=\{c_1, c_2\}} J(\{c_1, c_2\} | \{p_1\}, \{p_2\})$$

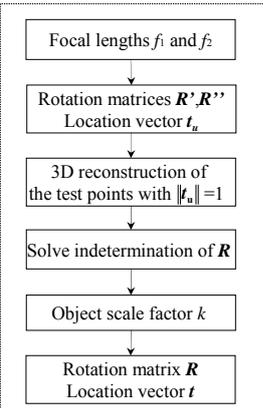
Elemento della popolazione: $\left\{ \mathbf{w} \left[\mathbf{c}_1(u_{c_1}, v_{c_1}), \mathbf{c}_2(u_{c_2}, v_{c_2}) \right]^T \right\}$



Evolution strategy

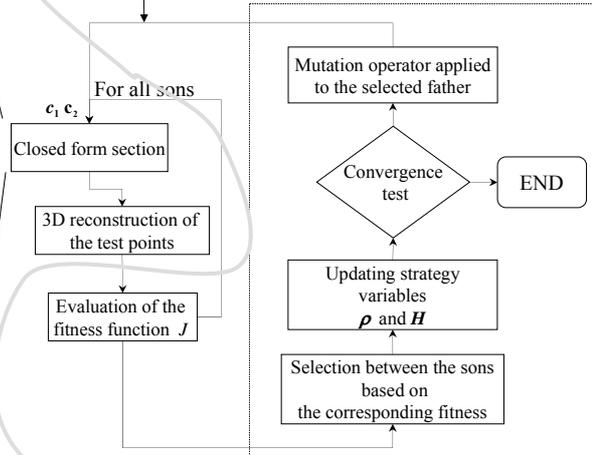


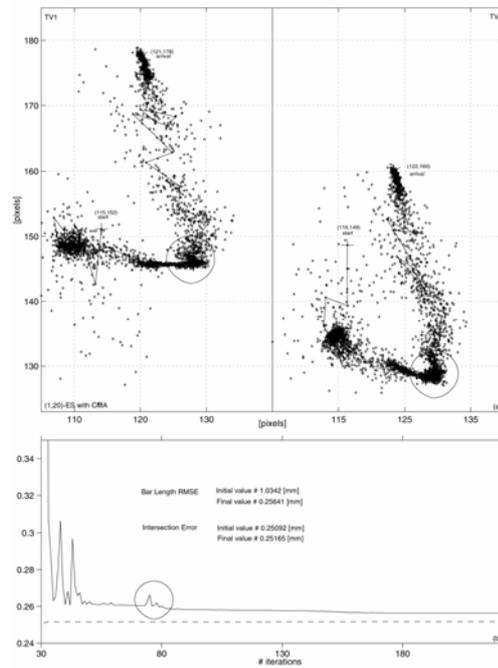
(a) Closed form section



Initialization from (M+λ)ES

(b) Iterative section





Dinamica
della
popolazione
delle
soluzioni

A.A. 2006-2007

homes.dsi.unimi.it/~borghese\



Progetti

Come progetto per il corso potete implementare una variante di strategia evolutiva da applicare a problemi reali (e.g. routing, scheduling, ...).

A.A. 2006-2007

54/55

http://homes.dsi.unimi.it/~borghese\



Sommario



- Algoritmi genetici
- Ottimizzazione evolutiva
- Applicazione delle strategie evolutive alla visione