

Capítol 4: Algorismes Genètics

Ref.

An Introduction to Genetic Algorithms

Melanie Mitchell

MIT Press, 1997.

Contingut

1. Introducció
2. Fonaments
3. Algorisme genètic bàsic
4. Comparació amb altres mètodes
5. Aplicacions
6. Temes avançats

1. Introducció

- Des del principi, els ordinadors han estat pensats per modelitzar el cervell humà, imitant l'aprenentatge i simulant l'evolució biològica.
- En els anys 80's, la motivació biològica va resurgir i com a conseqüència tenim les xarxes neurals, algorismes d'aprenentatge i la computació evolutiva.
- Els algorismes genètics són un exemple de computació evolutiva.

Malhauradament també s'han usat per calcular trajectòries de míssils i descifratge de codis militars...

L'aprenentatge automàtic es va enfocar des d'un altre punt de vista que l'estrictament simbòlic estudiat fins al moment.

1. Introducció

Computació evolutiva

- Entre el anys 50's i 60's s'estudien els sistemes evolutius com a eina d'optimitzar problemes d'enginyeria.
- La idea bàsica era evolucionar a partir d'una solució cap a d'altres que siquin millor usant operadors inspirats en la genètica i la selecció natural.
- Els biòlegs usaben els ordinadors per simular l'evolució en els seus experiments.
- 1960's: John Holland inventa els Algorismes Genètics.

La intenció de Holland, però, no era la de solucionar problemes concrets sinó la d'estudiar el fenòmen d'adaptació de la mateixa manera que succeeix en la natura i importar els mecanismes als ordinadors.

1. Introducció

L'atractiu de l'evolució

- Per a què usar el concepte d'evolució en la resolució de problemes?
 - Molts problemes requereixen d'una cerca en un espai de solucions molt gran
Elaboració de proteïnes, prediccions financeres.
 - Molts problemes necessiten ser resolts de forma adaptativa per poder funcionar sota canvis en l'entorn
Robots, interfícis d'usuari.
 - Alguns problemes requereixen programes que siguin innovatius
 - Molts problemes requereixen solucions complexes
Intel·ligència Artificial.

La determinació de l'estructura de les proteïnes és un problema que estudien molt les empreses farmacèutiques, per exemple. Es tracta de trobar la seqüència d'aminoàcids, entre totes les possibles, que compleix unes determinades propietats (com, per exemple, atacar un virus).

En un espai de solucions molt gran, el paral·lelisme pot ser útil si s'usa d'una forma efectiva. Cal, però, definir les estratègies per seleccionar el conjunt de solucions a explorar al mateix temps.

Ja veurem que també es poden fer prediccions financeres en xarxes neuronals. Aquí es tracta de cercar el conjunt de regles o equacions que modelitzen la predicció.

Com a programes innovadors entenem algorismes que realitzen una tasca computacional nova o un descobriment científic.

Les regles i els sistemes experts provenen d'un enfocament top-down de resolució de problemes. Però no sempre és possible elaborar regles per resoldre un problema. Altres enfocaments de tipus bottom-up basats en regles simples poden resultar en un comportament emergent per la aplicació paral·lela i interacció de les regles.

Entre els enfocaments bottom-up tenim els sistemes connexionistes (xarxes neuronals), els sistemes multi-agent i els algorismes genètics.

1. Introducció

L'atractiu de l'evolució

- L'evolució biològica proporciona un font d'inspiració per atacar aquests problemes
 - En la biologia hi ha tantes seqüències genètiques possibles com solucions a un problema, i només les solucions “desitjables” són les que sobreviuen.
 - Les espècies evolucionen a mesura que evolucionen les condicions ambientals sota les quals viuen
 - L'evolució també “dissenya” mecanismes innovatius, com és el cas dels gèrmens que invadeixen el cos d'un mamífer.
 - Les regles de l'evolució són molt simples:
 - Variació aleatòria (ex. mutació)
 - Selecció natural
 - Només l'espècie més adaptada sobreviu i es reproduïx.

A més de proporcionar tot l'espai de solucions possibles, l'evolució per si mateixa és un procés de cerca en paral·lel, ja que moltes espècies evolucionen al mateix moment.

2. Fonaments

Terminologia

Cellules:

- Tots els sers vius estan compostats per cel.lules.
- Cada cel.lula d'un ser viu conté el mateix conjunt de cromosomes

Cromosoma:

- Cadenes de DNA que serveixen de “marca d'identitat” d'un organisme.
- Constituit per gens

Gens:

- Blocs funcionals de DNA, cadascun dels quals codifica una proteïna concreta
- Codifica un tret característic

Els termes emprats estan inspirats en la biologia real, però s'usen d'un forma molt més simple.

Un gen codifica un tret característic d'una persona, com el color dels ulls.

2. Fonaments

Terminologia

Al·lel:

- Els possibles valors d'un tret característic

Locus:

- Posició que ocupa un gen en el cromosoma

Genoma:

- Col·lecció completa de cromosomes d'un organisme

Genotip:

- Conjunt de gens d'un genoma
- Dos individus amb el mateixos genoms tenen el mateix genotip

Per exemple, pel gen de color d'ulls els al·lels blau, castany, verd.

2. Fonaments

Terminologia

Organismes diploids:

- Els seus cromosomes estan aparellats.
- En la reproducció hi ha una recombinació o **creuament** de cromosomes

Organismes haploids:

- Els seus cromosomes no estan aparellats.
- Els gens s'intercanvien entre els dos pares

Gamet:

- Cromosoma simple provinent d'un organisme diploid

L'ésser humà és diploid, amb 23 parells de cromosomes.

En la reproducció dels organismes diploids,

- 1) els gens de cada pare s'intercanvien entre cromosomes de manera que es forma un gamet
- 2) els gamets de cada dos pares es combinen per crear un cromosoma diploid nou.

2. Fonaments

Terminologia

Mutació:

- Els descendents són susceptibles de patir mutacions
- L'objectiu de la mutació és corregir errors.

Fitness (adaptació)

- Probabilitat de que l'organisme es reproduïxi (viabilitat).
- Nombre de descendents (fertilitat)

Quan diem "corregir errors" ens referim a que hi ha una manca d'adaptació a l'entorn.

2. Fonaments

Interpretació i representació en els AG

Cromosoma:

- Solució candidata a un problema
- Normalment codificada com una cadena de bits

Gens

- Bit simple o subconjunt de bits que codifiquen un tret particular de la solució.

Al·lel

- 0 o 1 (si es una cadena de bits)

Genotip

- Configuració dels bits en un individu

És molt important un bon dimensionament dels cromosomes per tal de que els valors que codifiquin permetin abarcar tots els valors produïbles al problema a resoldre i donin la precisió necessària per arribar a trobar la solució, permetent a l'algorisme convergir més eficientment i donar una solució més acurada. Segons aquesta idea també cal pensar en l'efecte contrari, si no es desitja tanta precisió a la solució, amb pocs gens al cromosoma farà que l'algorisme genètic estalvi temps de procés innecessari i reduïm la complexitat del problema.

2. Fonaments

Interpretació algorísmica

Creuament (*crossover*):

- Intercanvi de bits entre dos cromosomes diploids

Mutació

- Canviar un bit en una posició (locus) aleatòria

2. Fonaments

Espais de cerca

- Col·lecció de solucions candidates a un problema
- Distància entre els diferents candidats
- Exemple: Disseny de proteïnes
 - L'espai de cerca és el conjunt de totes les proteïnes possibles
 - Suposem que limitem a les proteïnes de longitud 100
 - Hi ha 20 possibles aminoàcids per a cada posició
 - Codificació: cada aminoàcid → una lletra de l'alfabet
AGGMCGBL...
 - Distància: nombre de lletres diferents en les mateixes posicions
 $\text{distància}(\text{AGGMCGBL}, \text{MGGMCGBL}) = 1$
 $\text{distància}(\text{AGGMCGBL}, \text{LBMPAFGA}) = 8$

Quan diem proteïna de longitud 100, volem dir que estan constituïdes com a màxim per 100 aminoàcids.

2. Fonaments

Espai de cerca

- Un algorisme cerca una solució en l'espai de solucions
- **Neighboring** (veïnatge): els veïns a una solució tenen una qualitat similar → **Crossover**: els descendents d'una solució tindran una qualitat similar
- Un algorisme genètic processa poblacions de cromosomes, substituint una població per la següent.
- És necessari una funció d'adaptació (fitness o funció heurística) que avaluï l'adaptació de cada cromosoma
- *Fitness landscape*: Representació de l'espai de tots els possibles genotips amb la seva adaptació (*fitness*)

La funció d'adaptació, heurística o fitness avalua les aptituds dels individus enfront la solució on es vol arribar, és a dir, cal mesurar el grau de semblança o afinitat dels individus de la població amb els de la població final desitjada.

2. Fonaments

Fitness landscape

Figura 1.1

Per genotips de longitud $l=2$, on

$f(00)=0.7$

$f(01)=1.0$

$f(10)=0.1$

$f(11)=0.0$

En aquesta superfície podem trobar pics, valls, colines, etc.

L'evolució causa el moviment en aquesta superfície.

L'adaptació es pot interpretar com la tendència a moure's cap als pics.

(És una aproximació no realista des del punt de vista biològic ja que la funció d'adaptació d'un organisme no es pot separar de la resta d'organismes que comparteixen l'entorn. Però es fa servir en AG).

2. Fonaments

Exemple: optimització de funcions

- Es tracta de trobar els valors que fan òptima la funció:

$$f(y) = y + |\sin(32y)| \quad 0 \leq y \leq \pi$$

- Solucions candidates:
 - valors d'y
 - codificades en cadenes de bits que representen nombres reals
- Funcions d'adaptació
 - Tradueix una cadena de bits x, en un real y
 - Avaluat f(y)
 - f(y) és la funció d'adaptació de x

2. Fonaments

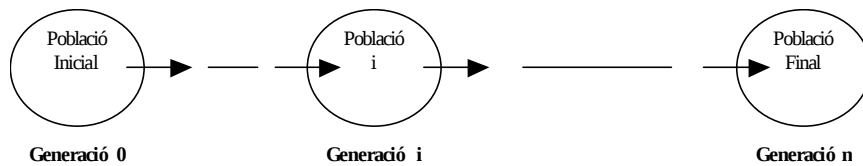
Exemple: disseny de proteïnes

- Problema: trobar seqüències de 50 aminoàcids que satisfan una determina estructura
- Hi ha 20 aminoàcids possibles, cadascun codificat per una lletra
- Cada proteïna està codificada per una cadena de 50 lletres
- Funció d'adaptació: negació de l'energia potencial respecte a l'estructura.
- Estimació de la funció: a través del càlcul de forces que actuen en cada aminoàcid

3. Algorisme genètic bàsic

Objectiu i funcionament

- L'objectiu és el de que generació a generació els individus estiguin més aprop de la solució del problema que es planteja
- El funcionament bàsic d'un algorisme genètic és el d'**evolucionar**, a partir d'una població inicial d'individus, cap a una altra població d'individus amb unes característiques desitjades.



3. Algorisme genètic bàsic

Propietats

- La **selecció** és indispensable per escollir els individus de la població més capacitats.
- El **creuament** aportarà variabilitat als individus de la nova generació.
- L'**herència** és necessària per transmetre les millores de les generacions pares a filles.
- La **mutació** permet en gran mesura el poder evolucionar cap a la solució desitjada quan la població es caracteritza per una manca d'aptituds, proporcionant-la de diversitat.

3. Algorisme genètic bàsic

Operadors

Els algorismes genètics es basen en tres tipus d'operadors bàsics:

- Selecció:
 - Selecciona els cromosomes per la reproducció
 - Amb un fitness més alts, el cromosoma pot ser escollit més cops
- Creuament (*crossover*):
 - Selecciona de forma aleatòria un locus i intercanvia les subsequències d'abans i després d'aquest locus entre dos cromosomes per crear un descendent.

10000100	→	10011111
11111111		11100100

- Mimetitza la reproducció diploid
- Un creuament succeïx amb una certa probabilitat (p_c)

3. Algorisme genètic bàsic

Operadors

- Mutació.
 - Altera de forma aleatòria un bit d'un cromosoma.
00000100 $\longrightarrow$ 01000100
 - Una mutació succeeix amb una probabilitat (p_m) molt petita (ex. 0.001)

3. Algorisme genètic bàsic

Generació d'una nova població

1. Inicialització aleatòria de la població amb n cromosomes de longitud l
2. Calcula el grau d'adaptació $f(x)$ de cada cromosoma x en la població.
3. Repetir, fins a generar n descendents:
 - a. Seleccionar un parell de cromosomes
 - b. Creuar els cromosomes d'acord amb la probabilitat p_c
Si no es produeix el creuament, els descendents són còpies exactes dels pares.
 - c. Mutar dos descendents amb probabilitat p_m
 - d. Posar els nous descendents en la nova població
 - e. Si n és senar, s'elimina de forma aleatòria un nou individu
4. Substituir la població per la nova
5. Anar al pas 2.

Dades:

- un problema
- una representació de solucions candidates basada en cadenes de bits

La selecció del cromosoma és basa en el valor de fitness. És una selecció "amb reemplaçament" en el sentit que un cromosoma pot ser escollit més d'un cop.

3. Algorisme genètic bàsic

- Aquest és l'algorisme que produeix una nova generació.
- La nova població estarà en terme mig més a prop de la solució que l'anterior.
 - ⇒ Repetint aquest procés és una de les principals raons perquè hi hagi **convergència** cap a la solució
- L'algorisme s'itera diverses vegades, i el conjunt resultant de la darrera generació s'anomena "**execució**".
- Degut a les probabilitats que intervenen, en cada execució s'obtingran resultats diferents.
- Per tant els resultats dels algorismes genètics es solen proporcionar en estadístiques o mitjanes d'execució.

Cal notar però que, si cap individu de la població inicial presenta una característica desitjada per la solució no es convergirà mai cap a aquesta. Així cal que hi hagi un grau de diversitat a cada població i això ho proporcionarà la mutació.

3. Algorisme genètic bàsic

Paràmetres

- Grandària de la població
- Longitud dels individus
- Nombre de generacions
- Probabilitat de creuament
- Probabilitat de mutació
- Funció d'adaptació

L'èxit de l'algorisme depèn en gran mesura dels paràmetres.

Cal que el nombre d'indivius de la població i el nombre de cromosomes dels individus sigui prou acurada. Cal ajustar aquests dos paràmetres el més alts possibles perquè l'algorisme convergeixi correctament a la solució i no caigui en màxims locals, si bé cal tenir en compte que augmentarà el temps de càlcul de l'algorisme i per tant el temps en trobar la solució. Un nombre d'individus prou alt assegurarà l'aparició de més esquemes o patrons genètics. Un nombre de gens prou alt permetrà codificar amb més precisió les dades que representen els gens i aportarà una solució més acurada. Normalment es treballa amb l'ordre de 100 individus de 20 a 30 gens per comosoma.

La probabilitat de creuament ha de ser prou alta perquè la mitja de valors objectiu s'estabilitzi després d'un nombre inicial de generacions. Els valors de treball adequats estan normalment entre el 75% i 95%.

La probabilitat de mutació ha de ser prou baixa perquè no afecti als bits més significatius en massa mesura. Els valors de treball adequats oscil·len entre el 0,5% i l'1%.

3. Algorisme genètic bàsic

Exemple. Definició i inicialització

- Longitud de les cadenes: $l=8$
- Funció d'adaptació: $f(x)$ = nombre de bits 1 en la cadena x
- Grandària de la població: $n=4$
- Probabilitats: $p_c=0.7$, $p_m=0.001$
- Inicialització (aleatòria):

Etiqueta	Cadena	adaptació	adaptació mitjana
A	00000110	2	12/4
B	11101110	6	
C	00100000	1	
D	00110100	3	

Exemple només com a il·lustració de l'algorisme.

3. Algorisme genètic bàsic

Exemple. Selecció

- Mètode de selecció: proporcional a l'adaptació

El nombre de vegades que un individu es reproduïx és proporcional a

$$\frac{f(x)}{\text{mitjana}(f(x_i))}$$

- Parelles seleccionades:

(B,D)

(B,C)

3. Algorisme genètic bàsic

Exemple. Creuament i mutació

- Creuament amb probabilitat p_c :
 - B i D es creuen en la primera posició --> descendents E, F
- Mutació amb probabilitat p_m :
 - E es muta en el locus 6: 10110100 → 10110000 (E')
 - F i C no es muten
 - B es muta en la posició 1: 11101110 → 01101110 (B')
- Nova població:

Etiqueta	Cadena	adaptació	adaptació mitjana
E'	10110000	3	14/4
F	01101110	5	
C	00100000	1	
B'	01101110	5	

4. Comparació amb altres mètodes de cerca

- Els AG són un mètode de cerca en l'espai de solucions que només exploren una petita fracció de tot l'espai.
- Els problemes de cerca de camins a objectius es resolten millor mitjançant mètodes de cerca en arbres.
- Els GA es consideren mètodes generals, “dèbils” (weak), com el hill climbing, simulated annealing i tabu search.
- Els GA es distingeixen d'altres algorismes de cerca per la seva combinació particular de
 - cerca paral·lela sobre una població, amb la
 - selecció estocàstica d'individus, creuament i mutació estocàstica.

En un mètode de cerca en arbres es poden avaluar en cada node solucions parcials, mentre que en un AG cal generar el conjunt complet de solucions. D'aquí que els AG no siguin adequats per resoldre problemes de camins, com per exemple, el nombre de moviments que calen en el problema del 8-puzzle.

5. Aplicacions

Dominis d'aplicacions

- Optimització: optimització numèrica, optimització combinatoria (circuits, job-shop scheduling)
- Programació automàtica: evolució de programes per a que realitzin determinades tasques, disseny d'autòmats, xarxes d'ordenació (*sorting networks*).
- Aprenentatge automàtic: classificació, predicció (del temps, de proteïnes), evolució de pesos en xarxes neurals, evolució de regles per sistemes de classificació, sensors per robots.
- Economia: modelització de processos d'innovació, desenvolupament d'estratègies d'ofertes, mercats emergents.

5. Aplicacions

Dominis d'aplicacions

- Ecologia: modelització de fenòmens ecològics com les carreres d'armaments biològics, evolució paràsits-hostes, simbiosi, flux de recursos.
- Genètica
- Sistemes socials: evolució social del comportament en les colònies d'insectes, evolució de la cooperació i comunicació en un **sistema multi-agent**.

5. Aplicacions

El dilema del presoner

- Dos individus (Àlícia i Bernat) són apresats, acusats d'haver comès un crim i engarjolats en dues cel·les separades, de forma que no es poden comunicar entre ells.
- A tots dos se'ls ofereix el següent:
 - Si acusa a l'altre, sortirà lliure i l'altre tindrà una condemna de 5 anys.
 - Si l'altre també l'acusa, tots dos tindran una condemna de 4 anys
 - Si cap acusa a l'altre, tindran una condemna de 2 anys.

⇒ Quina és l'estratègia millor per rebre la màxima recompensa?

El joc el va inventar Merrill Flood and Melvin Dresher als anys 50's.
Va intentar resoldre'l amb algorismes genètics Axelrod al 1987.

És un problema que s'ha estudiat molt en la teoria de jocs, en economia i en ciències polítiques perquè es pot veure com un model idealitzat de fenòmens reals com les carreres d'armes.

Àlícia ha d'acusar en Bernat arriscant-se a una condemna de 4 anys? Ha de cooperar confiant en Bernat per tenir tots dos una condemna de 2 anys? I si Bernat l'acusa?

Si sospites que el teu oponent cooperrà, el millor és acusar-lo.

Si sospites que t'acusarà, el millor és acusar-lo.

--> Sempre és millor acusar.

Però el dilema està en el fet que si els dos s'acusen, obtenen una recompensa pitjor que si tots dos cooperen.

Joc: cada jugador decideix de forma independent el moviment a realitzar.

5. Aplicacions

Dilema del presoner: Representació de les dades

1. Si cada jugador té memòria de la jugada prèvia, hi ha 4 casos:
CC, CD, DC, DD
 - Una estratègia és una regla que especifica l'acció a prendre en cadascun dels casos. En el TIT FOR TAT:
 - si CC, llavors C
 - si CD, llavors D
 - si DC, llavors C
 - si DD, llavors D
 - Que és pot representar en una cadena per CDCD.
2. Si cada jugador té memòria de 3 jugades anteriors, hi ha 64 casos:
CC CC CC, CC CC CD, CC CC DC, ..., DD DD DC, DD DD DD
 - La codificació d'una estratègia es pot fer mitjançant una cadena de 64 caràcters.
3. Finalment, es va usar una representació de 70 caràcters: 6 per expressar el cas, i 64 per expressar l'estratègia.

Robert Axelrod va sol·licitar a diferents investigadors que proporcionessin una estratègia. La millor va ser la de TIT FOR TAT:

- Cooperar en el primer joc
- En les successives jugades fer el que l'altre jugador ha fet en la jugada anterior.

Axelrod va implementar un AG per trobar la millor estratègia, que és el que es mostra aquí.

Per memòria d'una sola jugada:

Si A juga primer i B després: CD, llavors l'estratègia per A és D.

La grandària de l'espai de possibles solucions és, doncs, de 2^{70} (ja que hi ha 2 possibles valors per a cada posició de la cadena). És un espai molt gran per recórrer-lo exhaustivament.

C: cooperar

D: acusar

5. Aplicacions

El dilema del presoner: paràmetres i resultats

- Població inicial: 20 estratègies
- Funció d'adaptació: 3 experiments
 - Entorn fix
 - Es juga amb l'estratègia representada contra 8 estratègies generades per humans i el valor de puntuació més alt obtingut és la el valor d'adaptació.
 - Entorn canviant
 - Es juga amb l'estratègia contra cadascuna de les estratègies de la població.
 - Com la població evoluciona → l'entorn evoluciona.
- Resultats:
 - Exp 1: Es van obtenir estratègies més especialitzades que TIT-FOR-TAT, amb millors resultats
 - Exp 2: L'adaptació global de les estratègies va augmentar com a conseqüència de la cooperació (combat entre elles).

A més d'aquest resultats, es va provar la utilitat dels AG en l'obtenció de solucions evolutives.

5. Aplicacions

Xarxes d'ordenació

- Una xarxa d'ordenació és un dispositiu paral·lelitzable per ordenar llistes d'una grandària n d'elements.
- Problema: determinar el nombre mínim de comparacions necessàries per a ordenar correctament una xarxa de grandària n .

Les xarxes d'ordenació van estar definides per (Knuth, 1973).

Els algorismes genètics han estat aplicats en les xarxes d'ordenació per (Hillins, 1990).

Fig. 1.4

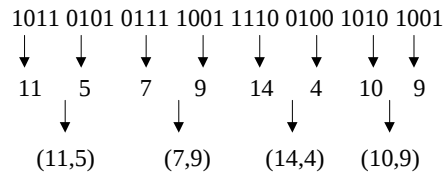
- Cada línia horitzontal representa un element de la llista
- Cada línia vertical representa una comparació a realitzar entre els dos elements
- Si els elements estan desordenats, s'intercanvien
- Comparacions en una mateixa columna es poden fer de forma paral·lela.

Observar que no es fa èmfasi en quines comparacions hauran de ser paral·leles, sinó que el nombre total de comparacions a realitzar sigui el mínim.

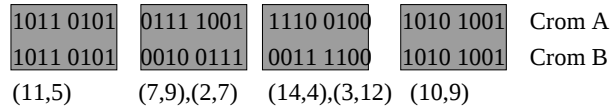
5. Aplicacions

Xarxes d'ordenació: representació de dades

- Cada cromosoma té 32 bits, cadascun representant una comparació



- La representació és diploide: cada individu consisteix en 15 parells de cromosomes:.

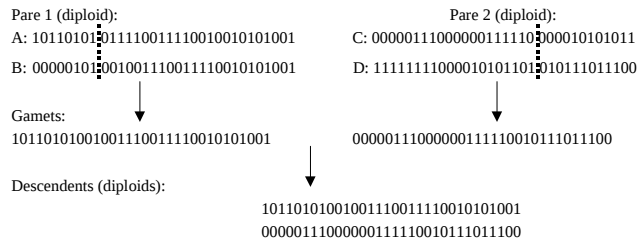


- Si un parell de cromosomes representen la mateixa comparació, aleshores es considera una sola comparació.
- ⇒ El nombre de comparacions pot oscil·lar entre 60-120.

5. Aplicacions

Xarxes d'ordenació: paràmetres

- La funció d'adaptació és el nombre de casos que ordena de forma correcta.
- Creuament:



- La probabilitat de mutació és de $p_m = 0.001$

5. Aplicacions

Xarxes d'ordenació: resultats

- Resultats inicials: L'AG va assolir mínims locals, amb solucions que no superaven les solucions proporcionades per humans.
 - Fenòmen de l'hoste-paràsit: els casos de test fagocitaven l'algorisme.
 - Solució: els paràsits (els casos de test) evolucionen juntament amb la població.
 - Resultats finals: no varen superar els proporcionats pels humans, però es van millorar molt.
- ⇒ Demostració dels AG com a tècniques per coevolució.

Coevolució: evolució dels paràsits juntament amb els hostes.