

Multi Expression Programming

Capitolul

16

- ❖ Structura cromozomilor
- ❖ Calitatea cromozomilor
- ❖ Efectul mutațiilor
- ❖ Efectul încrucișărilor
- ❖ Concluzii
- ❖ Rezumat
- ❖ Implementări sugerate

În cadrul acestui capitol vom introduce un nou tip de algoritmi genetici care vor opera asupra unor expresii și nu asupra unor valori. Vom prezenta structura unui cromozom, modalitatea prin care pot fi manipulați astfel de cromozomi și vom arăta că această structură este foarte eficientă în momentul în care se dorește folosirea algoritmilor genetici care operează asupra expresiilor aritmetice.

16.1. Structura cromozomilor

Fiecare genă a unui cromozom va reprezenta o subexpresie. Cel mai simplu tip de subexpresie constă într-o singură variabilă sau o constantă.

Celelalte subexpresii constau dintr-un operator și unul sau mai mulți operanzi (numărul acestora depinde de tipul operatorului). Operanzii sunt subexpresii corespunzătoare altor gene ale cromozomului.

Pentru a evita referințele ciclice este impusă următoarea structură: numărul de ordine (indicele) unei gene trebuie să fie mai mare decât numerele de ordine ale genelor care corespund operanzilor utilizați în cadrul genei respective.

Ca urmare, o genă constă dintr-un operator (de orice tip) și un număr de indici care identifică subexpresiile care reprezintă operanzii.

16.1.1. Construirea unui cromozom

Să presupunem că dorim să obținem un cromozom care să corespundă expresiei:

$$1 + 2 * (3 + 4) * 5 + (6 + 7) * 8.$$

Cromozomul trebuie să conțină cel puțin opt gene corespunzătoare celor opt valori numerice constante.

Pentru a identifica aceste gene vom introduce un operator special care operează asupra unui singur operand. Vom nota acest operator prin $@$, iar valoarea $@x$ va fi întotdeauna x .

Reprezentarea primelor opt gene ale cromozomului este prezentată în figura 16.1.

Gena #	Conținutul genei
1.	$@ 1$
2.	$@ 2$
3.	$@ 3$
4.	$@ 4$
5.	$@ 5$
6.	$@ 6$
7.	$@ 7$
8.	$@ 8$

Figura 16.1: Primele opt gene ale cromozomului care descrie expresia $1 + 2 * (3 + 4) * 5 + (6 + 7) * 8$

Se poate observa imediat faptul că subexpresia $3 + 4$ poate fi foarte ușor descrisă ca fiind operatorul $+$ aplicat asupra genelor #3 și #4. Același mecanism poate fi utilizat pentru a descrie expresia $6 + 7$.

Așadar, avem nevoie de două gene suplimentare, așa cum se poate vedea în figura 16.2.

Gena #	Conținutul genei
1.	$@ 1$
2.	$@ 2$
3.	$@ 3$
4.	$@ 4$
5.	$@ 5$
6.	$@ 6$
7.	$@ 7$
8.	$@ 8$
9.	$+ \#3 \#4$
10.	$+ \#6 \#7$

Figura 16.2: Primele zece gene ale cromozomului care descrie expresia $1 + 2 * (3 + 4) * 5 + (6 + 7) * 8$

Folosind notația $\#j$ pentru cea de-a j -a genă a cromozomului, expresia poate fi rescrisă astfel:

$$\#1 + \#2 * (\#3 + \#4) * \#5 + (\#6 + \#7) * \#8.$$

Folosind și nou-introdusele gene $\#9$ și $\#10$, expresia devine:

$$\#1 + \#2 * \#9 * \#5 + \#10 * \#8.$$

Vom introduce acum două noi gene care corespund operațiilor $\#2 * \#9$ și $\#10 * \#8$ (subexpresiile corespunzătoare acestor gene sunt $2 * (3 + 4)$ și $(6 + 7) * 8$). Prima dintre aceste două gene conține operatorul $*$, aplicat asupra genelor $\#10$ și $\#8$, în timp ce cea de-a doua conține același operator aplicat asupra genelor $\#2$ și $\#9$.

Se poate observa că aceste gene nu au fost introduse în ordinea "naturală". Aceasta se datorează unei foarte importante proprietăți a acestei structuri: ordinea a două gene independente nu este importantă.

În figura 16.3 este prezentată structura cromozomului în acest moment:

Gena #	Conținutul genei
1.	@ 1
2.	@ 2
3.	@ 3
4.	@ 4
5.	@ 5
6.	@ 6
7.	@ 7
8.	@ 8
9.	+ #3 #4
10.	+ #6 #7
11.	* #10 #8
12.	* #2 #9

Figura 16.3: Primele douăsprezece gene ale cromozomului care descrie expresia

$$1 + 2 * (3 + 4) * 5 + (6 + 7) * 8$$

Folosind aceste două noi gene, expresia devine:

$$\#1 + \#11 * \#5 + \#12.$$

Pentru a efectua operația $\#11 * \#5$ vom introduce o nouă genă care va avea forma $\#11 \#5$. În acest moment expresia devine:

$$\#1 + \#13 + \#12.$$

Adăugând gena $\#1 \#13$ obținem o expresie și mai simplă:

$$\#14 + \#12.$$

În final vom adăuga o ultimă genă care va avea forma + #14 #12. Această genă corespunde întregii expresii care trebuie descrisă de cromozom.

În concluzie, structura cromozomului care descrie expresia este cea prezentată în figura 16.4.

Gena #	Conținutul genei
1.	@ 1
2.	@ 2
3.	@ 3
4.	@ 4
5.	@ 5
6.	@ 6
7.	@ 7
8.	@ 8
9.	+ #3 #4
10.	+ #6 #7
11.	* #10 #8
12.	* #2 #9
13.	* #12 #5
14.	+ #1 #13
15.	+ #14 #12

Figura 16.4: Cromozomul care descrie expresia

$$1 + 2 * (3 + 4) * 5 + (6 + 7) * 8$$

16.1.2. Expresii corespunzătoare genelor

Așa cum am afirmat anterior, fiecare genă corespunde unei subexpresii. Unele gene sunt complet independente de altele, dar există și numeroase gene care depind una de alta.

De exemplu, genele simple (cele care conțin operatorul @) sunt întotdeauna independente unele de altele. De asemenea, se poate observa că și genele #9 și #11 sunt independente.

Genele #15 și #2 nu sunt independente deoarece gena #15 depinde de gena #12 (aceasta este unul din operanzii din gena #15), iar gena #12 depinde de gena #2.

Se poate spune că două gene sunt independente dacă o modificare în oricare dintre ele nu se reflectă în subexpresia corespunzătoare celeilalte.

Vom arăta acum că cromozomul prezentat anterior descrie în totalitate expresia luată în considerare. Vom construi subexpresiile corespunzătoare tuturor genelor și vom arăta că ultima genă corespunde întregii expresii.

În figura 16.5 sunt prezentate subexpresiile corespunzătoare pentru fiecare dintre cele cincisprezece gene.

Gena #	Conținutul genei	Subexpresia
1.	@ 1	1
2.	@ 2	2
3.	@ 3	3
4.	@ 4	4
5.	@ 5	5
6.	@ 6	6
7.	@ 7	7
8.	@ 8	8
9.	+ #3 #4	3 + 4
10.	+ #6 #7	6 + 7
11.	* #10 #8	(6 + 7) * 8
12.	* #2 #9	2 * (3 + 4)
13.	* #12 #5	2 * (3 + 4) * 5
14.	+ #1 #13	1 + 2 * (3 + 4) * 5
15.	+ #14 #12	1 + 2 * (3 + 4) * 5 + (6 + 7) * 8

Figura 16.5: Subexpresiile corespunzătoare genelor cromozomului din figura 16.4

16.1.3. Echivalențe

Datorită independenței unor gene, există o mulțime de cromozomi echivalenți. Un cromozom echivalent celui prezentat în figura 16.5 poate fi văzut în figura 16.6.

Gena #	Conținutul genei	Subexpresia
1.	@ 1	1
2.	@ 2	2
3.	@ 3	3
4.	@ 4	4
5.	@ 5	5
6.	@ 6	6
7.	@ 7	7
8.	@ 8	8
9.	+ #3 #4	3 + 4
10.	* #9 #5	(3 + 4) * 5
11.	* #2 #10	2 * (3 + 4) * 5
12.	+ #6 #7	6 + 7
13.	+ #1 #11	1 + 2 * (3 + 4) * 5
14.	* #12 #8	(6 + 7) * 8
15.	+ #13 #14	1 + 2 * (3 + 4) * 5 + (6 + 7) * 8

Figura 16.6: Un alt cromozom care descrie expresia $1 + 2 * (3 + 4) * 5 + (6 + 7) * 8$

Evident, nu este obligatoriu ca primele gene să conțină doar operatori @. Totuși, această abordare este mai simplă deoarece este mai ușor de stabilit o clasificare a genelor dacă se știe că variabilele și constantele sunt conținute de primele gene. Astfel, operațiile efectuate asupra celor două tipuri de gene sunt mai simplu de realizat.

Mai mult, rezultatele obținute folosind o astfel de repartizare a genelor nu sunt mai bune sau mai puțin bune decât cele obținute folosind un alt tip de repartizare.

16.1.4. Folosirea variabilelor

În cadrul cromozomilor prezentați anterior am utilizat doar constante pe post de operanzi. Structura este foarte simplă chiar dacă folosim variabile. În figura 16.7 este prezentată structura unui cromozom care caracterizează expresia $(x + y)^2$.

Gena #	Conținutul genei	Subexpresia
1.	@ x	x
2.	@ y	y
3.	+ #1 #2	$x + y$
4.	* #3 #3	$(x + y) * (x + y)$

Figura 16.7: Cromozomul care descrie expresia $(x + y)^2$

16.2. Calitatea cromozomilor

Unul dintre cele mai importante aspecte în cadrul utilizării algoritmilor genetici este calitatea unui cromozom, deoarece această caracteristică este utilizată în momentul în care sunt aleși cromozomii supraviețuitori.

Pentru un cromozom de tipul celui descris în acest capitol, putem stabili o calitate pentru fiecare dintre gene. Datorită faptului că fiecare genă corespunde unei expresii, teoretic, soluția problemei poate fi dată de oricare dintre genele cromozomului.

Ca urmare, o genă poate fi mai promițătoare decât o altă genă a unui cromozom, mai ales în cazul în care cele două gene sunt independente.

În concluzie, calitatea unui cromozom trebuie să depindă de calitățile individuale ale genelor din componența sa.

16.2.1. Calitatea genelor

În principiu, calitatea unei gene se determină folosindu-se o abordare similară celei utilizate în algoritmi genetici clasici. Așadar, calitatea trebuie să arate cât de aproape este soluția corespunzătoare genei față de soluția căutată.

Să presupunem că dorim să găsim o expresie care folosește valorile constante 2, 4, 5 și 7, operatorii “+” și “*”, iar valoarea expresiei să fie cât mai apropiată de 100. În timpul evoluției, putem obține un cromozom asemănător cu cel din figura 16.8.

Gena #	Conținutul genei	Subexpresia	Rezultatul
1.	@ 2	2	2
2.	@ 4	4	4
3.	@ 5	5	5
4.	@ 7	7	7
5.	+ #1 #2	$2 + 4$	6
6.	+ #2 #4	$4 + 7$	11
7.	* #5 #6	$(2 + 4) * (4 + 7)$	66
8.	+ #6 #6	$(4 + 7) + (4 + 7)$	22
9.	+ #3 #4	$5 + 7$	12
10.	* #2 #3	$4 * 5$	20
11.	* #6 #9	$(4 + 7) * (5 + 7)$	132
12.	+ #7 #8	$(2 + 4) * (4 + 7) + (4 + 7) + (4 + 7)$	88
13.	* #2 #3	$4 * 5$	20
14.	+ #12 #9	$(2 + 4) * (4 + 7) + (4 + 7) + (4 + 7) + (5 + 7)$	100
15.	+ #6 #8	$(4 + 7) + (4 + 7) + (4 + 7)$	33
16.	+ #15 #1	$(4 + 7) + (4 + 7) + (4 + 7) + 2$	35
17.	+ #7 #16	$(2 + 4) * (4 + 7) + (4 + 7) + (4 + 7) + (4 + 7) + 2$	101
18.	* #10 #3	$(4 * 5) * 5$	100
19.	+ #7 #15	$(2 + 4) * (4 + 7) + (4 + 7) + (4 + 7) + (4 + 7)$	99
20.	+ #8 #15	$(4 + 7) + (4 + 7) + (4 + 7) + (4 + 7) + (4 + 7)$	55

Figura 16.8: Un cromozom și rezultatele subexpresiilor

În acest moment trebuie să stabilim calitatea fiecărei gene; evident, există o sumedenie de variante. Cele mai simple dintre ele constau în determinarea erorii relative sau absolute a unei gene. Aceste valori vor arăta cât de aproape se află o expresie față de expresia căutată.

Se poate observa că cromozomul conține două rezultate perfecte: genele #14 și #17. Fiecare dintre ele reprezintă o soluție corectă, chiar dacă una dintre expresii este mai simplă.

De asemenea, se poate observa faptul că cromozomul conține două gene identice: #10 și #13. Chiar dacă acesta ar putea părea un aspect redundant al cromozomului, în următoarele generații cromozomul ar putea pierde una dintre aceste gene, iar expresia conținută de ele s-ar putea dovedi a fi foarte utilă.

Evident, rezultatele corespunzătoare anumitor gene sunt mai apropiate de soluția căutată decât rezultatele corespunzătoare altor gene. În figura 16.9 este prezentată calitatea genelor bazată pe eroarea absolută.

Gena #	Conținutul genei	Rezultatul	Calitatea
1.	@ 2	2	98
2.	@ 4	4	96
3.	@ 5	5	95
4.	@ 7	7	93
5.	+ #1 #2	6	94
6.	+ #2 #4	11	89
7.	* #5 #6	66	34
8.	+ #6 #6	22	78
9.	+ #3 #4	12	88
10.	* #2 #3	20	80
11.	* #6 #9	132	32
12.	+ #7 #8	88	12
13.	* #2 #3	20	80
14.	+ #12 #9	100	0
15.	+ #6 #8	33	67
16.	+ #15 #1	35	65
17.	+ #7 #16	101	1
18.	* #10 #3	100	0
19.	+ #7 #15	99	1
20.	+ #8 #15	55	45

Figura 16.9: Calitatea genelor cromozomului din figura 16.8

În acest caz, este foarte clar faptul că o calitate mai mică va corespunde unei gene mai performante (deși ar putea părea ciudat, se întâmplă frecvent în lumea algoritmilor genetici ca denumirea unei noțiuni să pară necorespunzătoare semnificației sale reale).

Modul în care se stabilește calitatea genelor depinde de problema a cărei rezolvare se dorește a fi găsită folosind această metodă.

16.2.2. Combinarea calităților genelor

Evident, un cromozom cu gene de calitate superioară este mai promițător decât un cromozom cu gene de calitate inferioară. Cu toate acestea, modalitatea exactă prin care se stabilește calitatea cromozomilor depinde de problema dată.

Cea mai simplă modalitate de a stabili calitatea cromozomului este de a considera că aceasta este egală cu calitatea celei mai performante gene. Această abordare este foarte ușor de implementat, dar are anumite limitări cauzate de faptul că doi cromozomi ar putea avea aceeași calitate (dacă cele mai promițătoare gene au aceeași calitate) cu toate că celelalte gene au calități complet diferite.

O altă situație neplăcută este cazul în care un cromozom conține o genă foarte performantă și o mulțime de gene neperformante. Un alt cromozom ar putea avea o mulțime de gene aproape la fel de performante, dar ar pierde "bătălia" într-un "duel" direct.

Ca urmare, modalitatea de stabilire a calității cromozomului pe baza calității genelor sale este o alegere delicată. Evident, depinde foarte mult de problemă, dar cele mai simple variante sunt reprezentate de:

- suma calităților genelor;
- calitatea medie a celor mai performante k gene;
- produsul calităților genelor;
- media dintre calitatea celei mai performante gene și a celei mai neperformante gene.

16.3. Efectul mutațiilor

În cadrul acestei secțiuni vom prezenta efectul puternic pe care îl poate avea o mutație asupra unui cromozom. În cazul în care acesta descrie o expresie în formatul *MEP* (*Multi Expression Programming*) introdus în cadrul acestui capitol, mutațiile sunt foarte importante datorită dependențelor dintre gene.

Ca de obicei, mutațiile contează la alterarea unei singure gene. De obicei, pentru cromozomii MEP nu este recomandată alterarea genelor simple (cele care conțin operatorul @).

Vom arăta acum cât de drastic poate fi modificată semnificația unui cromozom în urma unei mutații foarte simple.

Să considerăm din nou cromozomul din figura 16.8 și să presupunem că dorim să alterăm gena #6 prin modificarea conținutului său din + #2 #4 în + #2 #3.

Chiar dacă se modifică un singur indice, genele cromozomilor care depind de gena modificată au acum o semnificație complet diferită. Efectul mutației este prezentat în figura 16.10.

Gena #	Conținutul genei	Subexpresia	Rezultatul
1.	@ 2	2	2
2.	@ 4	4	4
3.	@ 5	5	5
4.	@ 7	7	7
5.	+ #1 #2	2 + 4	6
6.	+ #2 #3	4 + 5	9
7.	* #5 #6	(2 + 4) * (4 + 5)	54
8.	+ #6 #6	(4 + 5) + (4 + 5)	18
9.	+ #3 #4	5 + 7	12

10.	* #2 #3	$4 * 5$	20
11.	* #6 #9	$(4 + 5) * (5 + 7)$	108
12.	+ #7 #8	$(2 + 4) * (4 + 5) + (4 + 5) + (4 + 5)$	72
13.	* #2 #3	$4 * 5$	20
14.	+ #12 #9	$(2 + 4) * (4 + 5) + (4 + 5) + (4 + 5) + (5 + 7)$	84
15.	+ #6 #8	$(4 + 5) + (4 + 5) + (4 + 5)$	27
16.	+ #15 #1	$(4 + 5) + (4 + 5) + (4 + 5) + 2$	29
17.	+ #7 #16	$(2 + 4) * (4 + 5) + (4 + 5) + (4 + 5) + (4 + 5) + 2$	85
18.	* #10 #3	$(4 * 5) * 5$	100
19.	+ #7 #15	$(2 + 4) * (4 + 5) + (4 + 5) + (4 + 5) + (4 + 5)$	83
20.	+ #8 #15	$(4 + 5) + (4 + 5) + (4 + 5) + (4 + 5) + (4 + 5)$	45

Figura 16.10: O mutație efectuată asupra cromozomului din figura 16.8

Chiar dacă în memoria calculatorului s-au modificat doar câțiva biți, schimbarea se amplifică datorită dependențelor și semnificația cromozomului este acum complet alta.

16.4. Efectul încrucișărilor

Pentru cromozomii *MEP* încrucișarea se realizează într-o manieră asemănătoare celei care este utilizată pentru cromozomii clasici. La fel ca și în cazul mutațiilor, efectele pot fi drastice. În figura 16.11 este prezentată o încrucișare cu un punct de tăiere pentru doi cromozomi *MEP*. Tăierea s-a realizat după gena #5.

Gena #	Conținutul genei în primul părinte	Subexpresia pentru primul părinte	Rezultatul pentru primul părinte
1.	@ 2	2	2
2.	@ 4	4	4
3.	@ 5	5	5
4.	@ 7	7	7
5.	+ #1 #2	$2 + 4$	6
6.	+ #2 #3	$4 + 5$	9
7.	* #5 #6	$(2 + 4) * (4 + 5)$	54

Gena #	Conținutul genei în al doilea părinte	Subexpresia pentru al doilea părinte	Rezultatul pentru al doilea părinte
1.	@ 2	2	2
2.	@ 4	4	4
3.	@ 5	5	5

4.	@ 7	7	7
5.	+ #3 #4	$5 + 7$	12
6.	* #2 #3	$4 * 5$	20
7.	+ #4 #5	$7 + (5 + 7)$	19

Gena #	Conținutul genei în primul fiu	Subexpresia pentru primul fiu	Rezultatul pentru primul fiu
1.	@ 2	2	2
2.	@ 4	4	4
3.	@ 5	5	5
4.	@ 7	7	7
5.	+ #1 #2	$2 + 4$	6
6.	+ #2 #3	$4 + 5$	9
7.	+ #4 #5	$7 + (2 + 4)$	13

Gena #	Conținutul genei în al doilea fiu	Subexpresia pentru al doilea fiu	Rezultatul pentru al doilea fiu
1.	@ 2	2	2
2.	@ 4	4	4
3.	@ 5	5	5
4.	@ 7	7	7
5.	+ #3 #4	$5 + 7$	12
6.	* #2 #3	$4 * 5$	20
7.	* #5 #6	$(5 + 7) * (4 * 5)$	240

Figura 16.11: Încrucișarea cu un punct de tăiere pentru cromozomii *MEP*

16.5. Concluzii

Este destul de clar faptul că nu vor exista prea multe probleme în care soluția să fie dată de o expresie. Totuși *MEP* este foarte utilă în rezolvarea unor probleme.

Dacă este utilizată corect, *MEP* permite găsirea unor expresii care ar putea fi folosite pentru a rezolva anumite probleme. De exemplu, pentru un joc ar putea fi găsită o expresie care să ducă la obținerea celei mai bune mutări într-o configurație dată. O alternativă ar fi să se caute o expresie care să caracterizeze cât mai bine posibil cât de sigură este o configurație dată.

În concluzie, *MEP* se dovedește a fi un instrument foarte util, chiar dacă va fi folosit în majoritatea cazurilor doar pentru cercetări științifice. Există șanse destul de mari ca, folosind *MEP*, să descoperiți expresii care pot fi utilizate în rezolvarea unor pro-

bleme NP-complete și care duc la rezultate mult mai bune decât algoritmi aproximativi folosiți în prezent.

16.6. Rezumat

Acest capitol a fost dedicat unei noi modalități de a utiliza algoritmi genetici asupra unor expresii matematice. Rezultatele obținute folosind *Multi Expression Programming* sunt foarte promițătoare. De exemplu, *Mihai Oltean*, cel care a prezentat pentru prima dată *MEP*, a reușit să determine o funcție care poate fi utilizată cu succes pentru problema comis-voiajorului. El a găsit o funcție care stabilește continuarea drumului în funcție de ponderile muchiilor apropiate de nodul curent.

16.9. Implementări sugerate

Dacă doriți să vă testați capacitățile de a utiliza *MEP*, vă sugerăm să încercați să rezolvați următoarele probleme folosind această tehnică:

1. jocul Tic-Tac-Toe (X și 0) – încercați să identificați o formulă care să poată fi utilizată pentru a stabili care este cea mai bună mutare într-o configurație dată;
2. încercați să scrieți un algoritm *MEP* care încearcă să determine expresia unei funcții pentru care se cunosc valorile în anumite puncte;
3. problema comis-voiajorului.