

Υλοποίηση Γενετικού αλγορίθμου για το πρόβλημα του Mastermind (Παράδοση της εργασίας 11/04/10).

Το πρόβλημα του Mastermind προέρχεται από ένα πολύ γνωστό επιτραπέζιο παιχνίδι με το όνομα Mastermind, το οποίο επινοήθηκε το 1970-71 από τον Ισραηλινό ειδικό τηλεπικοινωνιών Mordecai Meirowitz και εν συνεχεία τα δικαιώματά του παιχνιδιού αγοράστηκαν από την Αγγλική εταιρία Invicta Plastics Ltd. Σήμερα το Mastermind ανήκει σαν εμπορική ονομασία στην Pressman Toy Corporation, σε συνεργασία με την Invicta Toys and Games, Ltd., UK.

Περιγραφή του Προβλήματος

Στο παιχνίδι συμμετέχουν δύο παίκτες, αυτός που «βάζει» τον κωδικό (code-maker ή hider) και αυτός που προσπαθεί να «σπάσει» τον κωδικό (code-breaker ή seeker). Ο κώδικας είναι μια σειρά από χρωματιστά καρφάκια. Ο code-breaker κάνει μια μαντεψιά και ο code-maker του απαντά με τη μορφή 2 αριθμών: του αριθμού από καρφάκια που συμφωνούν τόσο στο χρώμα όσο και στη θέση που είναι τοποθετημένα (στο παιχνίδι ο αριθμός αυτός συμβολίζεται με αντίστοιχο αριθμό από μαύρα καρφάκια) και του αριθμού από καρφάκια που συμφωνούν μόνο στο χρώμα αλλά διαφέρουν στη θέση (στο παιχνίδι ο αριθμός αυτός συμβολίζεται με αντίστοιχο αριθμό από άσπρα καρφάκια). Για να αποκτήσετε μια εικόνα του παιχνιδιού, αλλά και να εξοικειωθείτε με αυτό παίζοντάς επισκεφτείτε τη σελίδα του Παγκόσμιου Ιστού:

http://www.telemath.gr/mathematical_telegames/mathematical_telegames_mastermind/maths_telegames_mastermind.php.

Το Mastermind διακρίνεται στο απλό (ή συνηθισμένο) στο οποίο έχουμε 6 διαφορετικά χρώματα και 4 θέσεις όπου μπορούν να μουν τα χρωματιστά ξύλινα καρφάκια και στο super Mastermind όπου έχουμε 8 διαφορετικά χρώματα και 5 διαφορετικές θέσεις.

Έτσι, λοιπόν, το γενικευμένο πρόβλημα Mastermind μπορεί να τυποποιηθεί σαν την αναζήτηση για ένα κρυφό κώδικα, ο οποίος χρησιμοποιεί «υπαινιγμούς» (ανάδραση) που παράγονται από ένα μαύρο κουτί (ο code-maker). Ειδικότερα ο κώδικας και οι μαντεψιές εξάγονται από ένα αλφάβητο (το αλφάβητο των χρωμάτων) με πληθάρημο (cardinality) N (το πλήθος των διαφορετικών χρωμάτων). Κάθε μαντεψιά και ο αντίστοιχος κώδικας έχει μήκος L . Έτσι ο χώρος αναζήτησης, που προκύπτει, έχει μέγεθος. Για $N=8$ και $L=5$ έχουμε την περίπτωση του Super Mastermind, στο οποίο ο χώρος αναζήτησης έχει μέγεθος 32768, ενώ για $N=6$ και $L=4$ έχουμε το απλό Mastermind στο οποίο ο χώρος αναζήτησης έχει μέγεθος 1296.

Γενετικοί αλγόριθμοι και Mastermind

Εδώ θα ασχοληθούμε με μια απλοποιημένη εκδοχή του απλού Mastermind ($N=6$ και $L=4$):

Με ένα συμπαίκτη ή με αντίπαλο τον εαυτό σου χρησιμοποίησε ένα Γενετικό Αλγόριθμο για να μαντέψεις μια σειρά Mastermind, από χρωματιστά ξύλινα καρφάκια.

Ο ένας παίκτης (ο code-maker) καθορίζει μια σειρά από τέσσερα χρώματα (4) (από αριστερά προς τα δεξιά) από ένα σύνολο έξι (6) χρωμάτων τα οποία είναι: Κόκκινο (K), Πράσινο (Π), Μπλε (Μ), Γαλάζιο (Γ), Ροζ (Ρ) και Βυσσινί (Β) και ο άλλος (ο code-breaker) θα πρέπει να χρησιμοποιήσει ένα Γενετικό Αλγόριθμο για να μαντέψει αυτή τη σειρά των χρωματιστών καρφιών. Για κάθε μαντεψιά ο code-breaker θα λαμβάνει ως ανάδραση από τον code-maker 2 βαθμούς για κάθε χρωματιστό καρφί το οποίο συμφωνεί τόσο στο χρώμα όσο και στη θέση, 1 βαθμό για κάθε χρωματιστό

καρφί που συμφωνεί μόνο στο χρώμα και 0 βαθμούς για καρφιά με χρώματα που δεν ανήκουν στον κώδικα.

Ο Γενετικός Αλγόριθμος που υλοποιεί το πρόβλημα έχει ήδη και δίνεται στα αρχεία `sga_mastermind.c` και `sga_mastermind.h`. Μελετήστε προσεκτικά τον κώδικα που υλοποιείται στα αρχεία αυτά. Ο στόχος σας σε αυτό το ερώτημα είναι να υλοποιήσετε δύο διαφορετικούς τελεστές (τεχνικές) επιλογής. Και οι δύο αυτές μέθοδοι έχουν προταθεί προκειμένου να αντιμετωπίσουν το πρόβλημα της πρόωρης σύγκλισης που εμφανίζεται αρκετά συχνά όταν χρησιμοποιείται για τη διαδικασία της επιλογής η κλασική μέθοδος της ρουλέτας (Roulette Wheel Selection). Υπάρχει, μία πεποίθηση στην επιστημονική κοινότητα, ότι η πρόωρη σύγκλιση οφείλεται στην παρουσία στον πληθυσμό σούπερ μελών (super individuals), τα οποία είναι πολύ καλύτερα (έχουν δηλαδή πολύ μεγαλύτερη απόδοση) από τα υπόλοιπα μέλη του πληθυσμού και κατά συνέπεια η απόδοσή τους είναι πολύ μεγαλύτερη από τη μέση απόδοση του πληθυσμού. Αυτό έχει ως αποτέλεσμα τα μέλη αυτά να έχουν πολύ μεγαλύτερη πιθανότητα να επιβιώσουν και να αναπαραχθούν σε σχέση με τα υπόλοιπα μέλη του πληθυσμού και έτσι μέσα σε λίγες γενιές «κατακυριεύουν» τον πληθυσμό εξαφανίζοντας τα υπόλοιπα χρωμοσώματα, καταστρέφουν την γενετική ποικιλομορφία και οδηγούν σε πρόωρη σύγκλιση.

- Μία διαφορετική (σε σχέση με τη μέθοδο που χρησιμοποιείται στον υπάρχοντα κώδικα) και ιδιαίτερα αποτελεσματική μέθοδος επιλογής είναι η μέθοδος του Linear Rank Based Selection. Σύμφωνα με τη μέθοδο αυτή τα μέλη του πληθυσμού κατατάσσονται με βάση την απόδοσή τους (Fitness). Στη συνέχεια αποδίδονται πιθανότητες επιλογής σε αυτά με βάση τη σειρά κατάταξή τους (Rank). Έτσι, εάν υποθέσουμε ότι το καλύτερο μέλος του πληθυσμού κατατάσσεται πρώτο, τότε η πιθανότητα της επιλογής ενός μέλους του πληθυσμού υπολογίζεται ως εξής: $prob(rank) = q - ((rank - 1) \cdot r)$. Όπου $rank$ είναι η θέση κατάταξης του μέλους και τα q και r δύο παράμετροι οι οποίες παίρνουν τιμές ως εξής: $q \in [1/N, 2/N]$ και

$$q = r \cdot (N - 1) / 2 + 1 / N \Rightarrow r = \frac{2 \cdot (N \cdot q - 1)}{N \cdot (N - 1)}, \text{ όπου } N = POPSIZE \text{ είναι το μέγεθος του}$$

πληθυσμού. Θα πρέπει να σημειωθεί ότι για τα μέλη του πληθυσμού ισχύει:

$$\sum_{i=1}^N prob(i) = 1. \text{ Υλοποιήστε και εντάξτε στον υπάρχοντα κώδικα τη μέθοδο του}$$

Linear Rank Based Selection.

- Μία άλλη μέθοδος επιλογής είναι η Tournament Selection. Σύμφωνα με τη μέθοδο αυτή για μέγεθος τουρνουά *TournamentSize*, επιλέγονται τυχαία *TournamentSize* διαφορετικά μέλη του πληθυσμού και ο νικητής του τουρνουά (αυτό το μέλος μεταξύ των μελών του τουρνουά με τη μεγαλύτερη απόδοση) επιλέγεται για να επιβιώσει. Δηλαδή, σε αυτή την περίπτωση δεν γίνεται καμία χρήση των πιθανοτήτων επιλογής (δεν υπολογίζονται καθόλου). Αυτό, φυσικά, επαναλαμβάνεται σε κάθε γενιά *POPSIZE* φορές προκειμένου να δημιουργηθεί ο ενδιάμεσος πληθυσμός. Υλοποιήστε και εντάξτε στον υπάρχοντα κώδικα τη μέθοδο του Tournament Selection.
- Πειραματιστείτε με τις δύο αυτές μεθόδους και συμπληρώστε τους παρακάτω πίνακες. Προκειμένου να υπολογίσετε το μέσο αριθμό γενεών θα πρέπει να «τρέξετε» τον αλγόριθμο τουλάχιστον 10 φορές και να υπολογίσετε τη μέση τιμή του αριθμού των γενεών που απαιτούνται μέχρι την επίλυση του προβλήματος.

Linear Rank Based Selection

A/A	ΜΕΓΕΘΟΣ ΠΛΗΘΥΣΜΟΥ	ΠΙΘΑΝΟΤΗΤΑ ΔΙΑΣΤΑΥΡΩΣΗΣ	ΠΙΘΑΝΟΤΗΤΑ ΜΕΤΑΛΛΑΞΗΣ	PARAMETER Q	ΜΕΣΟΣ ΑΡΙΘΜΟΣ ΓΕΝΕΩΝ ΓΙΑ ΤΗΝ ΕΥΡΕΣΗ ΛΥΣΗΣ
1	10	0.2	0.05	0.15	
2	30	0.2	0.05	0.05	
3	50	0.2	0.05	0.035	
4	50	0.2	0.05	0.025	
5	50	0.2	0.05	0.03	
6	50	0.2	0.05	0.04	

Tournament Selection

A/A	ΜΕΓΕΘΟΣ ΠΛΗΘΥΣΜΟΥ	ΠΙΘΑΝΟΤΗΤΑ ΔΙΑΣΤΑΥΡΩΣΗΣ	ΠΙΘΑΝΟΤΗΤΑ ΜΕΤΑΛΛΑΞΗΣ	TOURNAMENT SIZE	ΜΕΣΟΣ ΑΡΙΘΜΟΣ ΓΕΝΕΩΝ ΓΙΑ ΤΗΝ ΕΥΡΕΣΗ ΛΥΣΗΣ
1	10	0.2	0.05	3	
2	30	0.2	0.05	3	
3	50	0.2	0.05	3	
4	50	0.2	0.05	2	
5	50	0.2	0.05	5	
6	50	0.2	0.05	10	