

Γενετικός Προγραμματισμός

Εισαγωγή

Κεντρικός στόχος της Τεχνητής Νοημοσύνης αποτελεί η ανάπτυξη μεθόδων και τεχνικών που θα καταστήσουν τους Ηλεκτρονικούς Υπολογιστές ικανούς να επιλύουν προβλήματα με αυτόματο τρόπο. Από τις πρώτες εργασίες του Turing το 1948 έχουν γίνει πολλά βήματα και απομένουν ακόμη πολλά να γίνουν για την επίτευξη του παραπάνω στόχου.

Ο Γενετικός Προγραμματισμός ή ΓΠ, (Genetic Programming ή GP) είναι μία τεχνική του Εξελικτικού Υπολογισμού που μπορεί να επιλύει αυτόματα προβλήματα χωρίς να απαιτεί από τον χρήστη – χειριστή να γνωρίζει ή να προσδιορίζει εκ των προτέρων τη μορφή ή τη δομή της λύσης. Γενικότερα ο ΓΠ είναι μία μέθοδος συστηματική ανεξάρτητη από το πεδίο ορισμού του προβλήματος, που επιτρέπει στους Η/Υ να επιλύουν προβλήματα με αυτοματοποιημένο τρόπο ξεκινώντας από μία δήλωση υψηλού επιπέδου του «τι πρέπει να γίνει» για να καταλήξει στη δημιουργία ενός προγράμματος που επιλύει το πρόβλημα.

Με τον όρο πρόγραμμα εννοούμε ένα ευρύ πεδίο αποτελεσμάτων που ξεκινάει από την δημιουργία ενός πραγματικού προγράμματος σε γλώσσα υψηλού επιπέδου ή τον σχεδιασμό ενός αλγορίθμου ή την εύρεση μίας συνάρτησης ή τη σύνθεση ενός κυκλώματος και μπορεί να καταλήξει στη σύνθεση μίας εικόνας ή μουσικής ή στην σύνθεση οδηγιών για την οδήγηση ενός ρομπότ. Το τι μορφή θα έχει το πρόγραμμα εξαρτάται από την αρχική δήλωση του προβλήματος δηλ «τι πρέπει να γίνει».

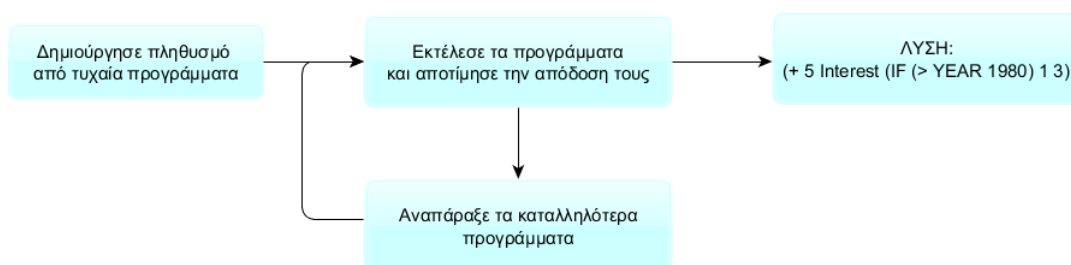
Το τελικό πρόγραμμα δηλ. η απάντηση του ΓΠ στο τιθέμενο πρόβλημα προκύπτει ως το εξαγόμενο μίας εξελικτικής διαδικασίας όπου εξελίσσεται ένας αρχικός πληθυσμός χιλιάδων τυχαίων προγραμμάτων σταδιακά και επιλέγεται ο καταλληλότερος μέσα από μία διαδικασία που χρησιμοποιεί τη Δαρβίνεια αρχή της φυσικής επιλογής, τη σχετική με την επιβίωση και αναπαραγωγή του καταλληλότερου.

Από την εμφάνιση του το 1992, ο ΓΠ έχει προσελκύσει το ενδιαφέρον χιλιάδων ανθρώπων σε παγκόσμιο επίπεδο. Διαρκώς σημειώνονται αλλαγές στο χώρο με την εμφάνιση καινοτόμων τεχνικών και νέων εφαρμογών από ερευνητές και επαγγελματίες, προσδίδοντας μία έντονη δυναμική στο χώρο στην προσπάθεια να απαντηθούν ερωτήματα, να βελτιωθεί και να επεκταθεί η μέθοδος.

Σύντομη περιγραφή του ΓΠ

Στον ΓΠ εξελίσσουμε έναν πληθυσμό από προγράμματα υπολογιστή. Επιδιώκουμε αυτή η εξέλιξη που λαμβάνει χώρα σε διακριτά χρονικά διαστήματα, τις γενεές, να γίνεται με ένα συνεχή ομαλό και στοχαστικό τρόπο, ελπίζοντας στην δημιουργία νέων καλύτερων πληθυσμών από προγράμματα όπως στην Εικόνα 1, αλλά όπως συμβαίνει στη φύση, η διαδικασία είναι τυχαία και δεν υπάρχει εγγυημένο αποτέλεσμα. Κι όλα αυτά, παρότι βασιζόμενος στην τυχειότητα ο ΓΠ μπορεί να οδηγηθεί σε διαφυγή από παγίδες (τοπικά μέγιστα) όταν άλλες ντετερμινιστικές μέθοδοι εγκλωβίζονται σε αυτά. Και παρόμοια με τη

φύση ο ΓΠ είναι σε θέση να εξελίξει νέες καινοτόμες και απρόσμενες λύσεις για τη λύση των τιθέμενων προβλημάτων.



Εικόνα 1: Η βασική ροή του ΓΠ, όπου η επιβίωση του καταλληλότερου χρησιμοποιείται για τη εύρεση της λύσης.

Τα βασικά βήματα του ΓΠ παρουσιάζονται στον παρακάτω *Αλγόριθμο 1*. Ο ΓΠ προσδιορίζει πόσο καλό είναι ένα πρόγραμμα εκτελώντας το και στη συνέχεια υπολογίζει τη συμπεριφορά του σε σχέση με μία ιδανική, πρότυπη τιμή που αποτελεί και στην ουσία τον προκαθορισμένο στόχο από τη δήλωση του προβλήματος, (Βήμα 2.1).

Για παράδειγμα, μας ενδιαφέρει αν το πρόγραμμα μας προβλέπει σωστά την τιμή του ρυθμού παραγωγής σε μία σειριακή γραμμή παραγωγής και τη συγκρίνει με την ιδεατή. Η απόκλιση των δύο τιμών ποσοτικοποιείται και γίνεται μία συγκεκριμένη τιμή που ονομάζεται καταλληλότητα (fitness) του προγράμματος. Τα προγράμματα πλέον έχοντας το κάθε ένα την τιμή της καταλληλότητάς του, μπορούν να καταταχθούν συγκριτικά μεταξύ τους μέσα στον πληθυσμό των προγραμμάτων και τα καλύτερα να αναπαραχθούν (Βήμα 2.2) και να δώσουν απογόνους δηλ νέα προγράμματα που θα δημιουργήσουν τον πληθυσμό της επόμενης γενεάς (Βήμα 2.3). Οι κυριότεροι γενετικοί τελεστές που χρησιμοποιούνται για την δημιουργία νέων προγραμμάτων είναι :

Διασταύρωση (crossover) : Από δύο επιλεγμένα γονικά προγράμματα συνδυάζονται τυχαία επιλεγμένα τμήματα που δημιουργούν το πρόγραμμα απόγονο.

Μετάλλαξη (mutation): Από ένα επιλεγμένο γονέα-πρόγραμμα δημιουργείται ένας απόγονος με την τυχαία διαμόρφωση ενός τυχαία επιλεγμένου τμήματος του γενετικού υλικού του γονικού προγράμματος.

Το γενετικό υλικό, δηλ. οι οδηγίες που έχουν κωδικοποιηθεί και απαρτίζουν κάποιο πρόγραμμα αποτελούν το **γονότυπο** του προγράμματος ενώ η καταλληλότητά του αποτελεί το **φαινότυπο**. Έτσι ο γενετικός προγραμματισμός κατατάσσεται στα συστήματα γονότυπου- φαινότυπου κατ' αναλογία με τους ζωντανούς οργανισμούς στη φύση.

-
1. Τυχαία δημιουργήσε έναν αρχικό πληθυσμό προγραμμάτων από τα διαθέσιμα αρχέτυπα
 2. **Επανάλαβε**
 1. Εκτέλεσε κάθε πρόγραμμα και υπολόγισε την καταλληλότητά του

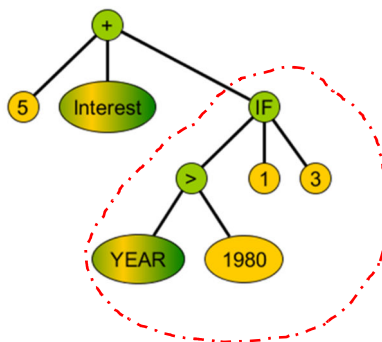
2. Επέλεξε ένα ή δύο προγράμματα από τον πληθυσμό με κάποιο πιθανοκρατικό τρόπο βασισμένο στην καταλληλότητα για να συμμετέχουν στις διαδικασίες αναπαραγωγής
 3. Δημιούργησε νέα προγράμματα εφαρμόζοντας γενετικούς τελεστές με προσδιορισμένες πιθανότητες
 4. **Αν δεν ικανοποιείται** το κριτήριο τερματισμού ή κάποια άλλη συνθήκη τερματισμού, επανέλαβε τα παραπάνω βήματα του βρόγχου
3. **Επέστρεψε** ως αποτέλεσμα τον καλύτερο ως τώρα άτομο

Αλγόριθμος 1

Αναπαράσταση στο ΓΠ

Ο συνηθέστερος τρόπος για την αναπαράσταση των προγραμμάτων στο ΓΠ είναι σαν **συντακτικά δένδρα (syntax trees)** αντί γραμμών κώδικα όπως θα περίμενε κάποιος. Ο λόγος χρήσης των δένδρων έχει να κάνει με την ευκολία που παρέχουν κατά την εκτέλεση των γενετικών τελεστών αλλά και της ευκολίας στην αποτίμηση. Παρακάτω στην Εικόνα 2 παρουσιάζεται το δένδρο του προγράμματος της Εικόνας 1.

Σε ένα δένδρο διακρίνουμε τον **ριζικό κόμβο (root)(+)** που είναι η αρχή του δένδρου, τους ενδιάμεσους εσωτερικούς κόμβους που καλούνται **συναρτήσεις (functions) (+,IF,>)** και τους εξωτερικούς κόμβους ή φύλλα του δένδρου που καλούνται **τερματικά (terminals)** και περιλαμβάνουν μεταβλητές και σταθερές (5, Interest, YEAR, 1980, 1, 3). Το σύνολο των επιτρεπόμενων κόμβων και τερματικών μαζί αποτελούν το **σύνολο αρχετύπων (primitive set)**. Σε ένα δένδρο ανάλογα με το που προσδιορίζουμε το ριζικό κόμβο μπορούμε να έχουμε υποσύνολα δενδριτικής μορφής που καλούνται **υποδένδρα (subtrees)**.

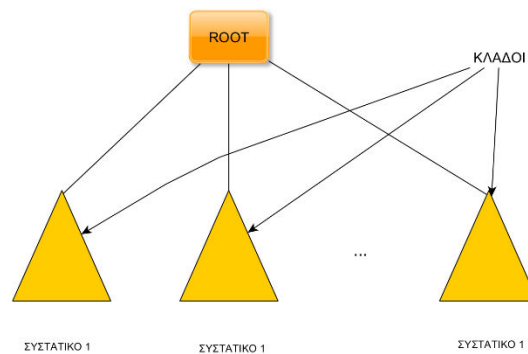


Εικόνα 2: Το δένδρο του προγράμματος της εικόνας 1. Διακρίνεται το υποδένδρο με διακεκομμένη εστισμένη γραμμή

Σε περισσότερο εξελιγμένες μορφές του ΓΠ, τα προγράμματα μπορεί να περιέχουν πολλαπλά συστατικά (πχ υπορουτίνες). Σε αυτή την περίπτωση αυτά αναπαρίστανται ως ένα σύνολο από δένδρα (ένα για κάθε ένα συστατικό) και ομαδοποιούνται κάτω από ένα ειδικό κόμβο που λειτουργεί σαν συνδετικό υλικό. Αυτά τα υποδένδρα καλούνται **κλάδοι**

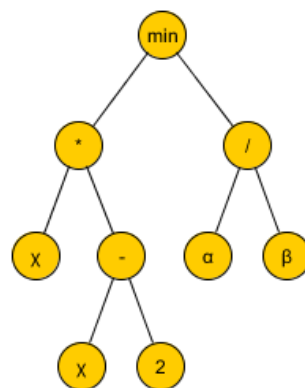
(branches). Ο αριθμός και ο τύπος των κλάδων σε ένα πρόγραμμα μαζί με κάποια άλλα χαρακτηριστικά της δομής τους απαρτίζουν την αρχιτεκτονική του προγράμματος (Εικ. 3).

Η πρώτη γλώσσα που χρησιμοποιήθηκε στις αρχές του ΓΠ ήταν η Lisp. Στη γλώσσα αυτή οι εντολές γραφόταν σε **προθεματική μορφή (prefix)** όπου ο τελεστής προηγείται των ορισμάτων π.χ. (+ 4 3 αντί της γραφής 4+3 **infix** που χρησιμοποιούμε). Αυτό το γεγονός διευκόλυνε την αναγνώριση των δένδρων διαμέσου των αντιστοιχιζόμενων σε αυτά εκφράσεων.



Εικόνα 3: Δένδρο με πολλαπλά συστατικά

Στα σύγχρονα περιβάλλοντα η αναπαράσταση των δένδρων στον Η/Υ έχει να κάνει κυρίως με το προγραμματιστικό περιβάλλον, γλώσσα προγραμματισμού και διαθέσιμες βιβλιοθήκες που χρησιμοποιούνται. Μία ευρεία γκάμα πλέον από γλώσσες προγραμματισμού και υπολογιστικά περιβάλλοντα χρησιμοποιούνται. C, C++,C#, Python , Ruby κοντά στις πρώτες Lisp και Prolog παράλληλα με τα εμπορικά πακέτα MatLab και Mathematica δίνουν πολλές επιλογές για ανάπτυξη κώδικα που υλοποιεί το ΓΠ. Πολλές φορές τα δένδρα δεν είναι ικανοποιητικά από πλευράς απόδοσης ή είναι επιθυμητά να χρησιμοποιούνται αρχέτυπα με πάνω από δύο ορίσματα **τάξης (arity)** 3 ή και παραπάνω. Ενδεικτικά η συνάρτηση AVG ή μέσος όρος, ας υποθέσουμε 5 ορισμάτων. Σε κάθε περίπτωση τα δένδρα μπορούν να μετατραπούν σε τάξης 2 (βλέπε Εικόνα 4) δηλ να μετατραπούν σε ένα υποδένδρο που κάθε κόμβος του θα έχει δύο μόνο ορίσματα ή άλλως δύο παιδιά.



Εικόνα 4: Δένδρο που αναπαριστά την $\min(x*x-2, a/b)$

Η χρήση δένδρων δεδομένης τάξης μπορεί να οδηγήσει σε αφαίρεση των παρενθέσεων χωρίς να υπάρξει παρανόηση. Έτσι το δένδρο της Εικόνας 4 γράφεται σε προθεματική μορφή ως: $(\min(* (x - x 2)) / \alpha \beta)$ ή με την αφαίρεση των παρενθέσεων : $\min * x - x 2 / \alpha \beta$.

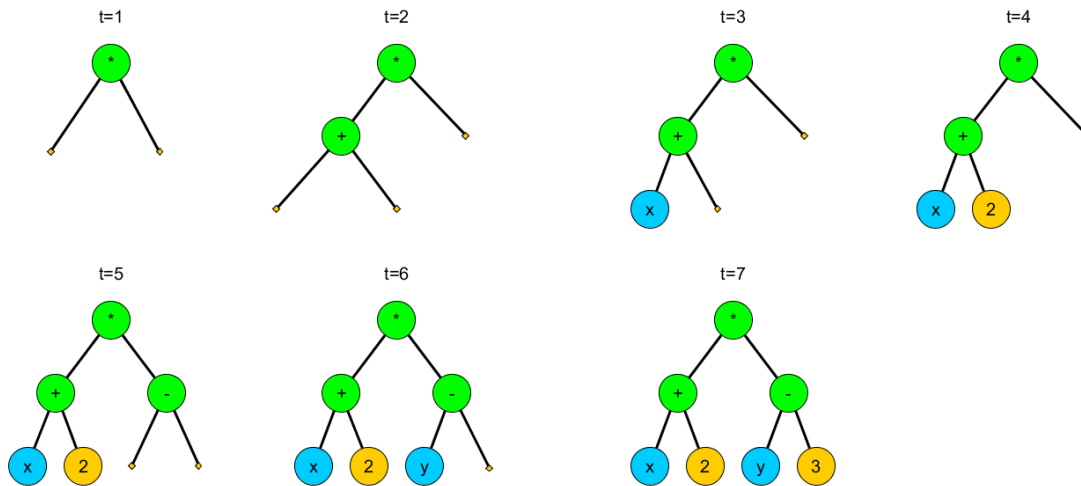
Αρχικοποίηση του πληθυσμού

Ένας αλγόριθμος ΓΠ ξεκινά την εκτέλεση του με τρόπο παρόμοιο όπως οι εξελικτικοί αλγόριθμοι. Τα άτομα που απαρτίζουν τον αρχικό πληθυσμό δημιουργούνται με τυχαίο τρόπο. Υπάρχουν διάφορες προσεγγίσεις για τον τρόπο δημιουργίας του πληθυσμού. Εδώ θα παρουσιαστούν δύο από τις παλαιότερες και απλούστερες μεθόδους για τη δημιουργία των ατόμων . Οι μέθοδοι **full** και **grow**.

Για να ξεκινήσουμε την δημιουργία των ατόμων πρώτα θα πρέπει να ορίσουμε το **βάθος (depth)** του δέντρου. Το βάθος είναι ο αριθμός των ακμών ξεκινώντας από το ριζικό κόμβο που έχει βάθος 0 έως έναν κόμβο που εξετάζουμε. Το βάθος του δένδρου είναι η απόσταση από το ριζικό κόμβο έως τον πλέον απομακρυσμένο κόμβο (φύλλο) του δένδρου.

Στη μέθοδο full, πήρε το όνομα της διότι παράγει πλήρη δένδρα δηλ. όλα τα φύλλα είναι στο ίδιο βάθος, οι κόμβοι παίρνονται τυχαία από το σύνολο συναρτήσεων έως ότου να φθάσουμε στο μέγιστο βάθος. Μετά από αυτό οι κόμβοι γεμίζονται από τυχαίες επιλογές από το σύνολο των τερματικών. Παρακάτω στην Εικόνα 5 παρουσιάζεται η διαδικασία δημιουργίας ενός τυχαίου δένδρου με τη μέθοδο full. Σαν σύνολο συναρτήσεων επιλέγεται $F=\{+,-,*,/\}$ και σαν σύνολο τερματικών $T=\{x,y,z,0,1,2,3,4,5\}$. Σημειώνουμε ότι όλες οι πράξεις του συνόλου συναρτήσεων έχουν τάξη (arity) ίση με 2 ενώ οι μεταβλητές (x,y,z) και οι σταθερές (0,1,2,3,4,5) έχουν τάξη ίσο με 0. Θα μπορούσαμε να έχουμε υπόψη, ότι η τάξη έχει τη σημασία του μέγιστου επιτρεπτού αριθμού απογόνων κόμβων σε ένα κόμβο γονέα.

Κατά τη στιγμή $t=1$ ένας κόμβος επιλέγεται τυχαία από το σύνολο συναρτήσεων σύμφωνα με την μέθοδο full. Αυτός έχει τάξη 2 άρα θα πρέπει να κρατηθούν δύο θέσεις για απογόνους κόμβους που θα γεμίσουν στα επόμενα στάδια. Την στιγμή $t=2$ ήδη έχουμε δύο ακμές βάθος οπότε δεν επιτρέπεται να προστεθεί άλλος κόμβος πλην τερματικός και έτσι επιλέγεται ο τερματικός κόμβος x την στιγμή $t=3$. Συνεχίζουμε με την ίδια λογική έως ότου να γεμίσουν όλοι οι κόμβοι.



Εικόνα 5: Δημιουργία δένδρου με τη μέθοδο full μέγιστου βάθους 2. Διαδοχικά στιγμιότυπα (t=time)

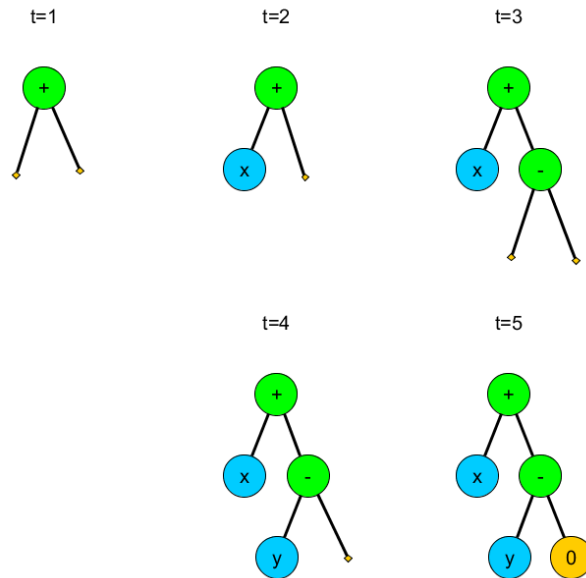
Πρακτικά υπάρχουν πολλοί τρόποι για να αναπαρασταθεί ένα δένδρο την μνήμη του Η/Υ. Στοιβές, διασυνδεδεμένες λίστες, συμβολοσειρές μεταξύ άλλων. Στον παρακάτω Πίνακα 1 παρουσιάζεται ο τρόπος αναπαράστασης της δημιουργίας του δένδρου με συμβολοσειρά. Το σύμβολο @ φανερώνει μία θέση κράτησης μνήμης που δημιουργείται ταυτόχρονα με την επιλογή ενός κόμβου από το σύνολο συναρτήσεων. Σε αυτή τη θέση θα εισαχθεί ένας απόγονος κόμβος, συνάρτηση ή τερματικό. Οι εκθέτες υποδηλώνουν τον κόμβο που έλκει την καταγωγή του ο κόμβος στη βάση του εκθέτη. Η τυχαία επιλογή των κόμβων γίνεται με κάποιο τρόπο από τη γεννήτρια ψευδοτυχαίων αριθμών. Θα πρέπει να δοθεί μέριμνα ώστε η επιλογή αυτή να είναι ομοιόμορφη ώστε κάθε μέλος του συνόλου των συναρτήσεων ή των τερματικών να έχει την ίδια πιθανότητα να επιλεγεί. Η ομοιόμορφη επιλογή τυχαίων αριθμών είναι ένα ζητούμενο που απαιτείται καθ' όλη την εκτέλεση του αλγορίθμου του γενετικού προγραμματισμού.

Πίνακας 1: Δημιουργία δένδρου σε συμβολοσειρά

Στιγμή	Θέσεις σε συμβολοσειρά
t=1	* @ * *
t=2	* + * @ + @ + @ *
t=3	* + * x + @ + @ *
t=4	* + * x + 2 + @ *
t=5	* + * x + 2 + * @ - @ -
t=6	* + * x + 2 + * y - @ -
t=7	* + * x + 2 + * y - 3 -

Η μέθοδος full παράγει δένδρα με τα φύλλα στο ίδιο βάθος χωρίς αυτό να σημαίνει ότι τα δένδρα θα έχουν το ίδιο **μήκος (length)** δηλ. αριθμό κόμβων. Αυτό γίνεται μόνον αν στο σύνολο συναρτήσεων όλα τα μέλη έχουν τον ίδιο αριθμό τάξης. Ακόμα όμως και τότε η **ποικιλομορφία (diversity)** δηλ. η παρουσία στον πληθυσμό δένδρων διαφορετικού σχήματος και μεγέθους είναι μικρή.

Η μέθοδος **grow** εξασφαλίζει μεγαλύτερη ποικιλία στη μορφή των παραγόμενων δένδρων επιτρέποντας σε κάθε επιλογή κόμβου αυτός να προέρχεται από το σύνολο των αρχετύπων (δηλ συναρτήσεις και τερματικά) έως ότου επιτευχθεί το μέγιστο βάθος. Από τη στιγμή εκείνη και μετά μόνο τερματικά επιτρέπονται.



Εικόνα 6 Δημιουργία δένδρου με τη μέθοδο grow

Στην Εικόνα 6 παρουσιάζεται η δημιουργία ενός δένδρου με μέγιστο βάθος 2. Το πρώτο όρισμα του ριζικού κόμβου είναι η μεταβλητή x , τάξεως μηδενικής, γεγονός που σηματοδοτεί την παύση της ανάπτυξης του αριστερού κλάδου του δένδρου αν και το μέγιστο βάθος δεν έχει επιτευχθεί. Στον παρακάτω Πίνακα 2 φαίνεται πως θα εξελιχθεί η αντίστοιχη συμβολοσειρά.

Στιγμή	Θέσεις σε συμβολοσειρά
t=1	$+@^+@^+$
t=2	$+x^+@^+$
t=3	$+x^+^-@^-@^-$
t=4	$+x^+^-y^-@^-$
t=5	$+x^+^-y^-0^-$

Τόσο η μέθοδος full όσο και η μέθοδος grow δεν εξασφαλίζουν την απαιτούμενη ποικιλία στα παραγόμενα δένδρα και για το λόγο αυτό προτάθηκε από τον Koza ο συνδυασμός των δύο παραπάνω μεθόδων που πρακτικά επιτυγχάνεται χρησιμοποιώντας ένα εύρος βάθους δένδρων αντί μίας μόνο τιμής (ramped) και παράγοντας τα μισά δένδρα με τη μέθοδο full και τα μισά με τη μέθοδο grow. Η συνδυαστική αυτή μέθοδος καλείται **ramped half-and-half**. Η μέθοδος αυτή δεν διαφοροποιεί τη μορφή των δένδρων όπως παρουσιάστηκαν παραπάνω αλλά επιδρά στην σύσταση του αρχικού πληθυσμού τροφοδοτώντας τον με δένδρα που εξασφαλίζουν μεγαλύτερη ποικιλομορφία στον πληθυσμό.

Επιλογή

Για την επιλογή των ατόμων του πληθυσμού που θα αναπαραχθούν, ο ΓΠ ακολουθεί την πρακτική που ακολουθείται και στους άλλους εξελικτικούς αλγορίθμους της πιθανοκρατικής επιλογής με βάση την καταλληλότητα του κάθε ατόμου. Με αυτόν τον τρόπο τα άτομα που έχουν υψηλή καταλληλότητα είναι περισσότερο πιθανό να αναπαραχθούν και να δημιουργήσουν προγράμματα απογόνους τους από ότι τα άτομα με χαμηλότερη καταλληλότητα. Η πλέον διαδεδομένη τεχνική για την επιλογή των απογόνων στο ΓΠ είναι η **επιλογή τουρνουά (tournament selection)** ενώ ακολουθεί η **αναλογική της καταλληλότητας επιλογή (fitness –proportionate selection)** χωρίς ωστόσο να αποκλείονται και άλλες τεχνικές επιλογής που χρησιμοποιούνται στους εξελικτικούς αλγόριθμους και που έχουν τεθεί σε χρήση ήδη από τα πρώτα χρόνια εφαρμογής των.

Στην επιλογή τουρνουά επιλέγεται ένας αριθμός από άτομα με τυχαίο τρόπο από τον πληθυσμό. Αυτά τα άτομα συγκροτούν ένα μικρό σύνολο και διαγωνίζονται μεταξύ τους με βάση την καταλληλότητα του κάθε ατόμου. Το άτομο που θα έχει την καλύτερη καταλληλότητα θα επιλεγεί σαν γονέας, ενώ στην περίπτωση που θα εκτελεστεί διασταύρωση και απαιτούνται δύο γονείς τότε θα εκτελεστεί η επιλογή τουρνουά δύο φορές. Αντίστοιχα ο επιλεκτικός μηχανισμός μπορεί να επιλέξει το χειρότερο από πλευράς καταλληλότητας άτομο και τη θέση του στον πληθυσμό να καταλάβει ένα νέο άτομο που προέρχεται από την εφαρμογή των γενετικών τελεστών σε άτομα που επιλέχθηκαν ως γονείς.

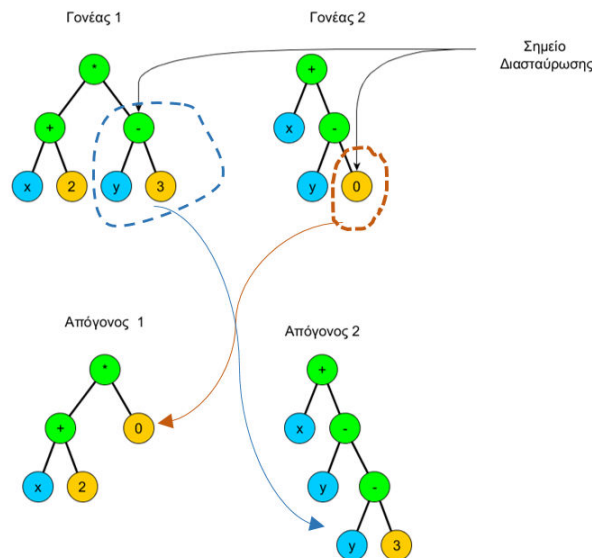
Η επιλογή τουρνουά αποτελεί ένα μηχανισμό για την εφαρμογή σταθερής επιλεκτικής πίεσης στον πληθυσμό. Αυτό συμβαίνει γιατί αρκεί ο νικητής να είναι κατά τι καλύτερος από τους άλλους χωρίς να εξετάζεται το πόσο καλύτερος είναι. Με αυτό τον τρόπο αποκλιμακώνεται η επιρροή της καταλληλότητας και αποφεύγεται η αντικατάσταση καλών προγραμμάτων από τους απογόνους τους, απευθείας, γεγονός που θα οδηγούσε στην μείωση γενιά με γενιά της ποικιλομορφίας, κατάσταση καταστροφική για την ικανότητα του πληθυσμού να παράγει λύσεις.

Η τυχαία επιλογή από τα άτομα του πληθυσμού δίνει τη δυνατότητα και σε μεσαίας καταλληλότητας προγράμματα να επιλεγούν και να αναπαραχθούν. Αυτό αποτελεί ένα πλεονέκτημα της μεθόδου που συνδυαζόμενο με την αποκλιμάκωση της καταλληλότητας και την ευκολία υλοποίησης της καθιστά τον συγκεκριμένο μηχανισμό επιλογής ιδιαίτερα ελκυστικό και διαδεδομένο στο χώρο του ΓΠ.

Γενετικοί Τελεστές

Ο ΓΠ διαφέρει σημαντικά σε σχέση με άλλους εξελικτικούς αλγορίθμους στο τρόπο που υλοποιεί τους γενετικούς τελεστές της διασταύρωσης και της μετάλλαξης. Αν και υπάρχουν αρκετοί τύποι τελεστή διασταύρωσης ο πλέον συχνά χρησιμοποιούμενος είναι η **διασταύρωση υποδένδρου (subtree crossover)**. Σε αυτόν τον τύπο δύο γονείς λαμβάνονται από το στάδιο της επιλογής και δημιουργούνται αντίγραφά τους. Αυτό γίνεται για να μην καταστραφούν τα πρωτότυπα των δύο γονέων τα οποία είναι το πιθανότερο προγράμματα υψηλής καταλληλότητας και είναι πιθανόν να επιλεγούν πολλές φορές για αναπαραγωγή. Σε κάθε ένα αντίγραφο ο τελεστής επιλέγει τυχαία και ανεξάρτητα ένα σημείο

διασταύρωσης δηλ έναν κόμβο σε κάθε δένδρο. Στη συνέχεια δημιουργείται (ή δημιουργούνται ανάλογα με την υλοποίηση) ένας (ή δύο) απόγονος με την αντικατάσταση του υποδένδρου με ριζικό κόμβο το σημείο διασταύρωσης στο αντίγραφο του 1^{ου} γονέα με το υποδένδρο που προέρχεται από το δεύτερο γονέα και έχει σαν ριζικό κόμβο το σημείο διασταύρωσης του 2^{ου} γονέα. Τα εναπομείναντα τμήματα ή διασταυρώνονται μεταξύ τους όπως στην Εικόνα 7, με αντίστοιχο τρόπο, είτε γίνονται σκουπίδια δηλ. αζήτητες περιοχές στην μνήμη του Η/Υ τις οποίες και χειρίζεται το σύστημα.



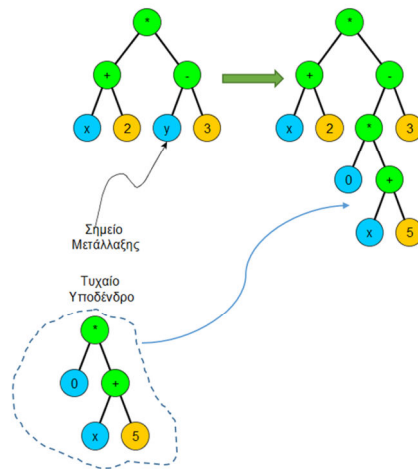
Εικόνα 7: Υλοποίηση Διασταύρωσης Υποδένδρου με δύο απογόνους (δεν χρησιμοποιείται συχνά). Οι γονείς είναι στην πραγματικότητα αντίγραφα των επιλεγέντων προγραμμάτων.

Συχνά για να λειτουργήσει αποτελεσματικά ο τελεστής τα σημεία διασταύρωσης επιλέγονται με μη ομοιόμορφο τρόπο. Τυπικά τα σύνολα των αρχετύπων του ΓΠ οδηγούν σε δέντρα με ένα μέσο **παράγοντα διακλάδωσης (branching factor)** τουλάχιστον ίσο με δύο. Ο παράγοντας διακλάδωσης είναι ο αριθμός των παιδιών που αντιστοιχούν σε κάθε κόμβο και η πλειοψηφία των κόμβων είναι φύλλα. Έτσι η ομοιόμορφη επιλογή σημείων διασταύρωσης οδηγεί τον τελεστή στο να παράγει μικρά δένδρα με αντίστοιχη μικρή ανταλλαγή γενετικού υλικού οπότε πολλές εφαρμογές του τελεστή εκφυλίζονται στην ανταλλαγή δύο φύλλων. Για την αποτροπή του παραπάνω ο Koza πρότεινε η εφαρμογή του τελεστή διασταύρωσης να γίνεται κατά 90% στους εσωτερικούς κόμβους, των συναρτήσεων και κατά 10% στους εξωτερικούς των τερματικών.

Η συνηθέστερος τύπος μετάλλαξης που χρησιμοποιείται στο ΓΠ είναι η **μετάλλαξη υποδένδρου (subtree mutation)** στην οποία ένας τυχαίος κόμβος επιλέγεται σαν σημείο μετάλλαξης που υποκαθίσταται από ένα τυχαία δημιουργούμενο υποδένδρο, Εικόνα 8.

Άλλος ένας συχνά χρησιμοποιούμενος τύπος μετάλλαξης είναι η **μετάλλαξη σημείου (point mutation)** βλ. Εικόνα 9, όπου ένας τυχαίος κόμβος επιλέγεται και το αρχέτυπο που περιέχει αντικαθίσταται τυχαία με ένα άλλο αρχέτυπο της ίδιας τάξης. Αν το τελευταίο δεν υπάρχει, δεν γίνεται καμία αλλαγή σε αυτόν τον κόμβο, αλλά οι άλλοι κόμβοι του δένδρου μπορεί να υφίστανται μετάλλαξη. Σε αυτό το σημείο έγκειται η βασική διαφορά του τελεστή από την μετάλλαξη υποδένδρου που συμβαίνει σε ένα δένδρο μία φορά σε αντίθεση με την μετάλλαξη σημείου που εφαρμόζεται ανά κόμβο. Με αυτό τον τρόπο κάθε κόμβος εξετάζεται και με μία ορισμένη πιθανότητα αλλάζει όπως αναφέρθηκε επιτρέποντας έτσι τη

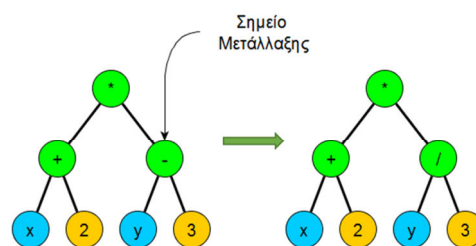
μετάλλαξη πολλών κόμβων ανεξάρτητα μεταξύ τους σε μία εφαρμογή του τελεστή σε ένα συγκεκριμένο δένδρο.



Εικόνα 8: Παράδειγμα μετάλλαξης υποδένδρου

Η επιλογή του ποιος τελεστή θα χρησιμοποιηθεί γίνεται πιθανοκρατικά με εξ αρχής ορισμένες πιθανότητες και κανονικά η εφαρμογή τους είναι αμοιβαία αλληλο-αποκλειόμενη. Η πιθανότητα του ποιος τελεστή θα χρησιμοποιηθεί ονομάζεται **αναλογία τελεστών (operator rates)** και συνήθως ο τελεστής διασταύρωσης απαντάται με μεγαλύτερο ποσοστό της τάξης του 90% αντίθετα με το πολύ χαμηλότερο ποσοστό της μετάλλαξης της τάξης του 1%.

Υπάρχει και ένας άλλος τελεστής αυτός της **αναπαραγωγής (reproduction)** που εφαρμόζεται με πιθανότητα $1-p$, όπου p το άθροισμα των πιθανοτήτων των τελεστών διασταύρωσης και μετάλλαξης. Ο τελεστής της αναπαραγωγής απλώς περιλαμβάνει την επιλογή των καταλληλότερων ατόμων και την και την εισαγωγή αντιγράφων τους στην επόμενη γενεά.



Εικόνα 9: Μετάλλαξη σημείου

Ενδεικτική Βιβλιογραφία

Γεωργούλη, Κ., 2015. *Τεχνητή νοημοσύνη. Μία εισαγωγική προσέγγιση*. Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών.

Λυκοθανάσης, Σ., 2001. *Γενετικοί Αλγόριθμοι και Εφαρμογές*. Πάτρα: Ελληνικό Ανοικτό Πανεπιστήμιο.

Τσάκωνας, Α. και Δούνιας, Γ., 2009. *Εξελικτικός υπολογισμός και εξόρυξη δεδομένων*. Αθήνα: Κλειδάριθμος.

Poli, R., Langdon, W.B. and McPhee, N.F., 2008. *A Field Guide to Genetic Programming*. Διαθέσιμο στη: <http://digitalcommons.morris.umn.edu/cs_facpubs/1>.