Η Μηχανική Μάθηση στο Σχολείο: Μια Προσέγγιση για την Εισαγωγή της Ενισχυτικής Μάθησης στην Τάξη

Σάββας Νικολαΐδης

1ο Γυμνάσιο Έδεσσας
savvas@sch.gr

Περίληψη

Η εργασία αυτή αποτελεί μια πρόταση για την εισαγωγή της Μηχανικής Μάθησης και συγκεκριμένα της Ενισχυτικής Μάθησης μέσω της μεθόδου Q-Learning στα σχολεία. Αναλύεται η μέθοδος και παρουσιάζεται γιατί η συγκεκριμένη μέθοδος αλλά και ο κλάδος της Μηχανικής Μάθησης γενικότερα είναι κατάλληλος για την εκπαιδευτική διαδικασία. Επίσης παρουσιάζεται η δυνατότητα να συνδυαστεί η μέθοδος Q-Learning με τη χρήση ρομπότ στο εκπαιδευτικό περιβάλλον.

Λέξεις κλειδιά: *Μηχανική Μάθηση, Ενισχυτική μάθηση, Q-Learning.*

Abstract

This paper is a proposal for the introduction of Machine Learning and more precisely the Q-Learning algorithm of the Reinforcement Learning method in schools. The method is analyzed and we present why this particular method and the whole Machine Learning field are relevant to the educational process. We also present the importance of the ability to connect the Q-Learning algorithm with the use of robots in the educational environment.

Keywords: *Machine Learning, reinforcement learning, Q-Learning*

1. Εισαγωγή

Τεχνητή Νοημοσύνη είναι ο κλάδος της επιστήμης υπολογιστών όπου προσπαθούμε να κάνουμε τους υπολογιστές να συμπεριφερθούν σαν νοήμονα όντα. Βασική λειτουργία των νοημόνων όντων είναι η μάθηση. Με τον όρο μάθηση εννοούμε την βελτίωση της συμπεριφοράς με την απόκτηση πείρας. Η μηχανική μάθηση είναι η περιοχή της τεχνητής νοημοσύνης όπου ασχολούμαστε με αλγορίθμους και μεθόδους που επιτρέπουν σε υπολογιστικά συστήματα να βελτιώνουν τη συμπεριφορά τους με την απόκτηση γνώσης. (Mitchell, 1997). Οι αλγόριθμοι μηχανικής μάθησης κατηγοριοποιούνται ανάλογα με το επιθυμητό αποτέλεσμα του αλγορίθμου. Οι συνηθέστερες κατηγορίες είναι οι εξής:

- **Μάθηση με Επίβλεψη** (Supervised Learning). Είναι η κλασική προσέγγιση του προβλήματος της μηχανικής μάθησης. Ο αλγόριθμος κατασκευάζει μια συνάρτηση που απεικονίζει δεδομένες εισόδους σε γνωστές, επιθυμητές

εξόδους (σύνολο εκπαίδευσης, υποδείγματα). Ο στόχος είναι η γενίκευση της συνάρτησης και για εισόδους με άγνωστη έξοδο.

- **Μάθηση χωρίς Επίβλεψη** (Unsupervised Learning). Ο αλγόριθμος κατασκευάζει ένα μοντέλο για κάποιο σύνολο εισόδων χωρίς να γνωρίζει επιθυμητές εξόδους για το σύνολο εκπαίδευσης.
- **Ενισχυτική Μάθηση** (Reinforcement Learning). Εδώ το σύστημα μας λειτουργεί με σκοπό να πετύχει έναν στόχο. Σαν καθοδήγηση στην αναζήτηση αυτή δεν έχει έναν άμεσο οδηγό όπως τα υποδείγματα στη μάθηση με επίβλεψη αλλά καθοδηγείται από "ανταμοιβές" και "τιμωρίες" που δέχεται ανάλογα με το αν οι ενέργειες του το φέρνουν πιο κοντά στη λύση ή όχι (Sutton & Barto, 1998). Σκοπός είναι να μεγιστοποιηθούν οι ανταμοιβές μακροπρόθεσμα. Τα αποτελέσματα των ενεργειών φαίνονται στο τέλος της διαδικασίας και γι αυτό η μέθοδος αυτή ονομάζεται και "μάθηση με καθυστέρηση" (Delayed Learning). Λειτουργεί σε καταστάσεις που ο στόχος είναι μακροπρόθεσμος σε ευμετάβλητα περιβάλλοντα και σε περιβάλλοντα για τα οποία έχουμε περιορισμένες πληροφορίες. Εφαρμογές βρίσκει στη ρομποτική, στα παιχνίδια κ.α.

2. Μηχανική Μάθηση στην τάξη

2.1 Ένταξη στο πρόγραμμα σπουδών και μαθησιακοί στόχοι

Το αντικείμενο της Μηχανικής Μάθησης μπορεί να ενταχθεί στο πρόγραμμα σπουδών του Γυμνασίου στην ενότητα "Προγραμματίζω τον Υπολογιστή" του άξονα μαθησιακών στόχων "Διερευνώ, ανακαλύπτω και λύνω προβλήματα με ΤΠΕ" η οποία σύμφωνα με το *"Πρόγραμμα Σπουδών για τον Πληροφορικό Γραμματισμό στο Γυμνάσιο"* διδάσκεται και στις τρεις τάξεις του Γυμνασίου.
Στην Α' Γυμνασίου στα πλαίσια της ενότητας αυτής μπορούν να παρουσιαστούν από τον εκπαιδευτικό έτοιμα παραδείγματα λογισμικού Μηχανικής Μάθησης όπου οι μαθητές θα κληθούν να αλλάξουν τη συμπεριφορά του συστήματος μεταβάλλοντας παραμέτρους. Στην Β' Γυμνασίου οι μαθητές καλούνται να κατανοήσουν τον αλγόριθμο Q-Learning που θα οδηγήσει έναν πράκτορα (agent) στην έξοδο λαβύρινθου, ενώ στην Γ' τάξη μπορούν να τον υλοποιήσουν προγραμματιστικά
Επίσης στην ενότητα "Υλοποιώ σχέδια έρευνας με ΤΠΕ" του άξονα μαθησιακών στόχων "Διερευνώ, ανακαλύπτω και λύνω προβλήματα με ΤΠΕ" η μέθοδος Q-Learning μπορεί να ενταχτεί στη δραστηριότητα της "Εκπαιδευτικής Ρομποτικής" όπου μπορούμε να εκπαιδεύουμε το ρομπότ να βρίσκει την έξοδο από έναν λαβύρινθο.
Η γλώσσας προγραμματισμού που χρησιμοποιείται στο σχολείο στα πλαίσια της διδασκαλίας της ενότητας "Προγραμματίζω τον Υπολογιστή" είναι ικανό εργαλείο για να υλοποιηθούν οι απαραίτητοι αλγόριθμοι. Απλοποιημένες μορφές της μεθόδου

Q-Learning μπορούν να υλοποιηθούν ακόμη και στο Microsoft Excel.
Ο βασικός στόχος της εισαγωγής της Μηχανικής Μάθησης είναι η εξοικείωση των μαθητών με έννοιες της τεχνητής νοημοσύνης. Οι μαθητές μέσα από αυτή τη προσέγγιση θα συνειδητοποιήσουν ότι η Τεχνητή Νοημοσύνη δεν είναι υποχρεωτικά κάτι το πολύπλοκο και δυσνόητο με το οποίο ασχολούνται μόνο οι επιστήμονες της πληροφορικής άλλα κάτι που το συναντάμε σε πολλές εφαρμογές της καθημερινότητάς μας, όπως για παράδειγμα τα παιχνίδια του υπολογιστή, καθώς επίσης και κάτι που θα μπορούσαν και οι ίδιοι σε κάποιο βαθμό να το υλοποιήσουν.

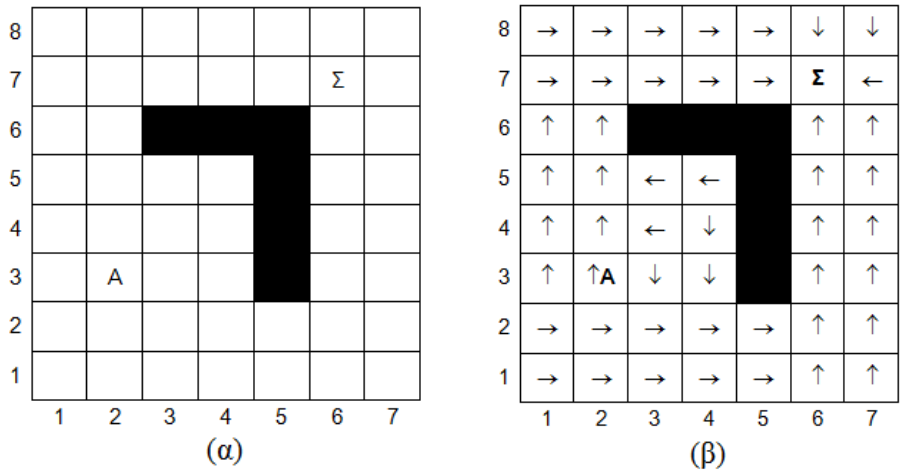
2.2 Η μέθοδος Q-Learning

Από τις διάφορες μεθόδους Μηχανικής Μάθησης πιστεύουμε ότι η μέθοδος της Ενισχυτικής Μάθησης και πιο συγκεκριμένα η μεθοδολογία Q-learning είναι αυτή που προσφέρεται καλύτερα για εφαρμογή στη τάξη. Οι λόγοι:

- Είναι εύκολα κατανοητή διαδικασία. Οι μαθητές μπορούν να κατανοήσουν τη λειτουργία της μεθόδου χωρίς πολλές προαπαιτούμενες γνώσεις.
- Κρατάει ζωντανό το ενδιαφέρον καθώς παρακολουθεί την "πρόοδο" του συστήματος που εκπαιδεύεται
- Η διαδικασία της εκπαίδευσης είναι αλληλεπιδραστική με δυνατότητα αλλαγών στο πεδίο όπως για παράδειγμα διαφορετική σχεδίαση του λαβύρινθου
- Οι μαθητές μπορούν να επέμβουν στην διαδικασία μάθησης ρυθμίζοντας παραμέτρους του συστήματος, όπως την παράμετρο λήθης, ώστε να επιτευχθούν τα καλύτερα δυνατά αποτελέσματα
- Προσομοιώνονται νοήμονες οργανισμοί και γίνονται πιο κατανοητές οι λειτουργίες τους.
- Μπορεί να εφαρμοστεί σε ρομπότ με φυσική υπόσταση. Έτσι υπάρχει αίσθηση ότι αυτό που γίνεται στον υπολογιστή έχει αντίκτυπο στον πραγματικό κόσμο.

Ας φανταστούμε έναν πράκτορα (agent) που έχει δυνατότητα να "κινείται" μέσα σε ένα συγκεκριμένο χώρο, ας πούμε έναν λαβύρινθο. Έχει μια κάποια αντίληψη του χώρου όσο του επιτρέπουν οι αισθητήρες του. Ας υποθέσουμε επίσης ότι αρχικά δεν έχει καμιά ιδέα για τις συνέπειες των πράξεων του, δηλαδή δεν ξέρει με ποιο τρόπο οι ενέργειες του θα αλλάξουν τα προϊόντα των αισθήσεών του. Αυτό που δέχεται είναι "αμοιβή" αν βρει πχ την έξοδο, και "τιμωρία" αν για παράδειγμα συγκρουστεί με κάποιο εμπόδιο. Έστω ότι ο πράκτοράς μας έχει δυνατότητα να κινείται στον *"κόσμο"* του *σχήματος 1α*. Κάθε στιγμή η θέση του καθορίζεται από τις συντεταγμένες του κελιού του. Ξεκινάει από το κελί Α με συντεταγμένες (2,3) και σε κάθε διακριτή μονάδα του χρόνου έχει τη δυνατότητα να κινηθεί μια θέση πάνω,

κάτω, δεξιά ή αριστερά. Στόχος του είναι να φτάσει στο κελί Σ με συντεταγμένες (6,7). Κάθε φορά που πέφτει σε τοίχο παίρνει "ανταμοιβή" -1 και όταν φτάσει στο Σ ανταμείβεται με +10. Όταν φτάνει στο Σ και αφού πάρει την ανταμοιβή του μεταφέρεται σε κάποιο άλλο τυχαίο κελί και συνεχίζει.



Σχήμα 1*: (α) ο "Κόσμος" και (β) η "Βέλτιστη Πολιτική"*

Αυτό που πρέπει τελικά να αποκτήσει ο πράκτοράς μας είναι μια *πολιτική*, συγκεκριμένες δηλαδή οδηγίες για το τι θα πρέπει να κάνει σε κάθε δεδομένη κατάσταση (σε κάθε κελί) για να πλησιάσει τον τελικό του στόχο. Η βέλτιστη πολιτική για τον πράκτορά μας φαίνεται στο *σχήμα 1β*. Το ερώτημα είναι πως θα καταφέρει ο πράκτορας να αποκτήσει την πολιτική αυτή με μόνο δεδομένο την συνάρτηση ανταμοιβής. Στη μάθηση με ενίσχυση είναι λογικό ότι ανταμοιβές στο κοντινό μέλλον έχουν μεγαλύτερη αξία από αμοιβές σε πιο απομακρυσμένη χρονική στιγμή. Για παράδειγμα θα προτιμήσουμε μια κίνηση που θα μας οδηγήσει στο στόχο σε 5 βήματα από μια άλλη που θα κάνει το ίδιο σε 10. Ορίζουμε λοιπόν έναν *συντελεστή λήθης γ* στο διάστημα [0,1].

Η μέθοδος Q-Learning προτάθηκε από τον Watkins, (Watkins & Dayan, 1992). Σύμφωνα με αυτή τη μέθοδο διατηρούμε έναν πίνακα τιμών Q(X,α) για κάθε ζευγάρι κατάστασης X και δυνατής δράσης α που αναπαριστά την ποιότητα της συγκεκριμένης δράσης στη συγκεκριμένη θέση. Η δράση που τελικά επιλέγεται και εκτελείται είναι αυτή για την οποία η Q δίνει τη μέγιστη τιμή. Η τιμή της Q(X,α) επηρεάζεται από την απευθείας αμοιβή για την κίνηση αυτή που δίνει η συνάρτηση αμοιβής r(X,α) καθώς και από τη μέγιστη τιμή της Q στη νέα θέση X' ως εξής:

$$Q(X,\alpha) \xleftarrow{\beta} r(X,\alpha) + \gamma \cdot V(X') \qquad \textbf{(1)}$$

όπου V(X') η μέγιστη τιμή του Q για τη νέα θέση, γ ο συντελεστής λήθης και β ο συντελεστής που μεταβάλλεται το Q από την παλιά στη νέα τιμή.

***Πίνακας 1**: Οι τιμές του Q*

X	**α**	**Q(X,α)**	**r(X,α)**
(2,3)	Π	4	0
(2,3)	Κ	6	0
(2,3)	Δ	3	0
(2,3)	Α	7	0
(1,3)	Π	5	0
(1,3)	Κ	4	0
(1,3)	Δ	2	0
(1,3)	Α	4	-1
…	…	…	…

Στο παράδειγμά μας ξεκινάμε με την Q να έχει τυχαίες τιμές σαν αυτές που φαίνονται στον πίνακα 1. Ξεκινώντας τον πράκτορα από τη θέση (2,3) βλέπει ότι η κίνηση α=Α προς τα αριστερά έχει το μέγιστο Q και έτσι μετακινείται στο (1,3) χωρίς να πάρει καμία ανταμοιβή από τη συνάρτηση r. Η μέγιστη τιμή στη νέα θέση είναι 5 για την κίνηση Π. Το σύστημα φέρνει τη τιμή της προηγούμενης κίνησης πιο κοντά σ' αυτή τη τιμή. Το Q για τη προηγούμενη κίνηση ήταν 7 και επομένως με συντελεστές γ=0.9 και β=2 έχουμε:

$$Q((2,3),A) = \frac{(0 + 0.9 \cdot 5) + 7}{2} = 5.75 \qquad \textbf{(2)}$$

και αντικαθιστούμε τη νέα τιμή για το Q στον πίνακα. Με τον ίδιο τρόπο συνεχίζουμε μέχρι το σύστημα να συγκλίνει σε μια καλή πολιτική. Το ότι βρήκαμε μια λύση για το πρόβλημα μας δεν σημαίνει απαραίτητα ότι βρήκαμε και την καλύτερη δυνατή λύση. Ο αλγόριθμος μας πρέπει να προβλέπει κάποτε και μερικές τυχαίες κινήσεις, έξω από την πολιτική της στιγμής, ούτως ώστε να έχουμε τη δυνατότητα να ξεφύγουμε από τη πεπατημένη και να πέσουμε πάνω στην καλύτερη λύση, αν βέβαια αυτή υπάρχει. Αυτές οι τυχαίες κινήσεις ονομάζονται *εξερευνητικές κινήσεις* σε αντίθεση με τις υπόλοιπες που τις αποκαλούμε *κινήσεις εκμετάλλευσης* (της ήδη αποκτημένης γνώσης). Το ποια θα είναι η σχέση ανάμεσα στην εξερεύνηση και την εκμετάλλευση και πως ακριβώς θα γίνουν οι εξερευνητικές κινήσεις εξαρτάται από το πρόβλημα που έχουμε να αντιμετωπίσουμε και τις συγκεκριμένες συνθήκες.

Βιβλιογραφία

Mitchell, T. (1997). *Machine Learning*, McGraw Hill.

Sutton R.S. & Barto A.G. (1998), *Reinforcement Learning*, MIT Press

Watkins and Dayan, C.J.C.H., (1992), *'Q-learning.Machine Learning'*, ISBN : 8:279-292

Διαθεματικό Ενιαίο Πλαίσιο Προγράμματος Σπουδών Πληροφορικής

Πρόγραμμα Σπουδών για τον Πληροφορικό Γραμματισμό στο Γυμνάσιο, 4η Έκδοση, Μάιος 2011