ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

UNIVERSITY OF THE AEGEAN

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

Μηχανική και Ενισχυτική Μάθηση για τα Συστήματα Συστάσεων

Φοιτήτρια:
Αμαλία-Μαρία Γιαννακοπούλου

Υπεύθυνος Καθηγητής:
Παναγιώτης Συμεωνίδης

Περιεχόμενα

Περιεχόμενα **i**

1 Περίληψη **1**

2 Εισαγωγή στη μηχανική μάθηση (Machine Learning) **3**

2.1 Τρόποι Μάθησης 4

2.1.1 Επιτηρούμενη μάθηση 4

2.1.2 Μη επιτηρούμενη μάθηση 5

2.1.3 Ενισχυτική μάθηση 5

3 Εισαγωγή στα νευρωνικά δίκτυα (Neural Network) **6**

3.1 Ιστορική αναδρομή 6

3.2 Ορισμός και Λειτουργία 7

3.3 Μοντέλο Τεχνητού νευρώνα 8

3.4 Συναρτήσεις Ενεργοποίησης (Activations Functions) 9

3.5 Αρχιτεκτονική Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ) 12

3.6 Προβλήματα στην αναγνώριση αντικειμένων 15

3.7 Κατηγορίες τεχνητών νευρωνικών δικτύων 16

3.7.1 Ο Αισθητήρας Perceptron 16

3.7.2 Δίκτυα Εμπρόσθιας Τροφοδότησης Ενός Επιπέδου . . . 18

3.7.3 Δίκτυα Εμπρόσθιας Τροφοδότησης Πολλών Επιπέδων (MLP) 19

3.7.4 Επαναλαμβανόμενο Νευρωνικό Δίκτυο (RNN) 22

3.7.5 Συνελικτικά Νευρωνικά Δίκτυα (CNN) 25
3.8 Δίκτυα Μακράς Βραχύχρονης Μνήμης(LSTM) 28
3.9 Εφαρμογές των τεχνητών νευρωνικών δικτύων 29

4 Ενισχυτική Μάθηση (Reinforcement Learning) 31
4.1 Εισαγωγή και Βασικά Στοιχεία 31
4.2 MDP (Markov Decision Process) 34
4.3 Μελέτη αλγορίθμου Q-learning 35
4.4 Μελέτη αλγορίθμου Deep Q-learning Network (DQN) 44
4.5 Μελέτη αλγορίθμου Advantage Actor-Critic learning (A2C) . . 49

5 Συστήματα Συστάσεων (Recommender Systems) 53
5.1 Εισαγωγή 53
5.2 Διαδικασία Σύστασης 54
5.3 Κατηγορίες Βασικών Μοντέλων Συστημάτων Συστάσεων . . . 54
5.3.1 Συστήματα Βασισμένα στη Συνεργασία (Collaborative Systems) 55
5.3.2 Συστήματα Συστάσεων Βάσει Περιεχομένου (Content-Based Recommender Systems) 57
5.3.3 Συστήματα Συστάσεων Βασισμένα Στη Γνώση (Knowledge-Based Recommender Systems) 59
5.4 Συστήματα Συστάσεων με MLP 61
5.4.1 Αλγόριθμος NeuMF (Neural Matrix Factorization) . . 62
5.4.2 Μαθηματική εξίσωση του NeuMF (Neural Matrix Factorization) 63
5.5 Συστήματα Συστάσεων με RNN 65
5.5.1 Αλγόριθμος GRU4Rec (Gated Recurrent Unit for Recommender Systems) 66
5.5.2 Μαθηματική εξίσωση του GRU4Rec (Gated Recurrent Unit for Recommender Systems) 67
5.6 Συστήματα Συστάσεων με A2C 68
5.6.1 Διαδικασία εκπαίδευσης και δοκιμής 70

5.6.2 Αλγόριθμος REINFORCE 72
5.6.3 Μαθηματική εξίσωση του Reinforce 73
5.7 Συστήματα Συστάσεων με DQN 74
5.8 Συστήματα Συστάσεων με CNN 74

6 Επίλογος **78**

Βιβλιογραφία **81**

Κεφάλαιο 1

Περίληψη

Η παρούσα διπλωματική εργασία επικεντρώνεται στην ολοκληρωμένη εφαρμογή της μηχανικής μάθησης και της ενισχυτικής μάθησης με στόχο τη βελτίωση των συστημάτων συστάσεων. Τα συστήματα συστάσεων αντιπροσωπεύουν ένα σημαντικό πεδίο έρευνας στον τομέα της τεχνητής νοημοσύνης και της πληροφορικής και αποσκοπούν στο να προβλέπουν και να παρέχουν εξατομικευμένες συστάσεις σε χρήστες βάσει των προτιμήσεών τους.

Η μηχανική μάθηση αποτελεί μια προσέγγιση που επιτρέπει στα συστήματα συστάσεων να αυτοματοποιήσουν τη διαδικασία της εξαγωγής πληροφορίας από τα δεδομένα και να προβλέπουν συστάσεις με βάση τα προηγούμενα μοτίβα και συμπεριφορές των χρηστών.

Από την άλλη πλευρά, η ενισχυτική μάθηση αναλαμβάνει να βελτιστοποιήσει την απόδοση των συστημάτων συστάσεων μέσω της αλληλεπίδρασης με το περιβάλλον και της ανταμοιβής για τις ενέργειες που πραγματοποιούνται. Μέσω αυτής της διαδικασίας, επιτυγχάνονται βελτιωμένες προτάσεις και πραγματοποιείται προσαρμογή του συστήματος στις μεταβαλλόμενες προτιμήσεις και ανάγκες των χρηστών.

Abstract

This dissertation focuses on the integrated application of machine learning and reinforcement learning to improve recommendation systems. Recommendation systems represent a significant research field in the areas of artificial intelligence and computer science, aiming to predict and provide personalized recommendations to users based on their preferences.

Machine learning is an approach that allows recommendation systems to automate the process of extracting information from data and predicting recommendations based on users' previous patterns and behaviors.

On the other hand, reinforcement learning takes on the task of optimizing the performance of recommendation systems through interaction with the environment and rewarding actions taken. Through this process, improved recommendations are achieved, and the system adapts to the changing preferences and needs of users.

Κεφάλαιο 2

Εισαγωγή στη μηχανική μάθηση (Machine Learning)

Η μηχανική μάθηση είναι ένας τύπος τεχνητής νοημοσύνης που δίνει την δυνατότητα στις εφαρμογές λογισμικού να προβλέπουν αποτελέσματα με ακρίβεια.Στην μηχανική μάθηση χρησιμοποιούνται αλγόριθμοι που προσπαθούν να προβλέψουν τιμές εξόδου χρησιμοποιώντας ιστορικά δεδομένα ως είσοδο.Επομένως, η δημιουργία μοντέλων ή προτύπων από ένα σύνολο δεδομένων ή από ένα υπολογιστικό σύστημα, ονομάζεται μηχανική μάθηση (machine learning).

Η Μηχανική Μάθηση έχει ως σκοπό τη δημιουργία μηχανών, ικανών να μαθαίνουν και να βελτιώνουν την απόδοσή τους σε κάποιους τομείς μέσω της αξιοποίησης προηγούμενης γνώσης και εμπειρίας.Για να επιτευχθεί το παραπάνω πρέπει να μελετηθούν και να κατασκευαστούν αλγόριθμοι που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά.Έτσι, δίνεται η δυνατότητα στις μηχανές να μαθαίνουν χωρίς να έχουν προγραμματιστεί ρητά.

Παραδείγματα εφαρμογών της μηχανικής μάθησης αποτελούν τα φίλτρα spam (spam filtering), η οπτική αναγνώριση χαρακτήρων (OCR), οι μηχανές αναζήτησης και η υπολογιστική όραση.

2.1 Τρόποι Μάθησης

Ανάλογα με τη φύση του προβλήματος, ο τομέας της Μηχανικής Μάθησης αναπτύσσει τρεις τρόπους μάθησης:

Figure 2.1: Τρόποι Μηχανικής Μάθησης
"Πηγή: https://ikee.lib.auth.gr/record/306657/files/GRI-2019-25332.pdf"

2.1.1 Επιτηρούμενη μάθηση

Η επιτηρούμενη μάθηση ή αλλιώς επιβλεπόμενη μάθηση (supervised learning) είναι ένας από τους τρόπους της μηχανικής μάθησης.Σε αυτό τον τρόπο το υπολογιστικό πρόγραμμα δέχεται τις εισόδους καθώς και τα επιθυμητά αποτελέσματα από έναν "δάσκαλο", και ο στόχος είναι να μάθει έναν γενικό κανόνα προκειμένου να αντιστοιχίσει τις εισόδους με τα αποτελέσματα.

'Εχει ονομαστεί έτσι καθώς υπάρχει κάποιος που επιβλέπει την σωστή τιμή εξόδου της συνάρτησης για τα δεδομένα που εξετάζονται.Τα δεδομένα κατανέμονται με βάση κάποια κριτήρια που ονομάζονται κλάσεις(Classes).

Η επιτηρούμενη μάθηση χρησιμοποιείται σε προβλήματα:

- Ταξινόμησης (Classification)
- Πρόγνωσης (Prediction)
- Διερμηνείας (Interpretation)

2.1.2 Μη επιτηρούμενη μάθηση

Η μη επιτηρούμενη μάθηση ή αλλιώς μάθηση χωρίς επίβλεψη (unsupervised learning) πρόκειται για τον δεύτερο τρόπο μηχανικής μάθησης.Η κύρια διαφορά με την επιτηρούμενη μάθηση είναι ότι σε αυτόν τον τύπο δεν παρέχονται οι είσοδοι.Επομένως, δεν υπάρχει κάποια εμπειρία στον αλγόριθμο μάθησης με αποτέλεσμα να πρέπει να βρεθεί η δομή των δεδομένων εισόδου.[1]

Χρησιμοποιείται σε προβλήματα:

- Ανάλυσης Συσχετισμών (Association Analysis)
- Ομαδοποίησης (Clustering)

2.1.3 Ενισχυτική μάθηση

Η ενισχυτική μάθηση (Reinforcement Learning) προσπαθεί να μάθει μέσα από την άμεση αλληλεπίδραση με το περιβάλλον για να πετύχει τον στόχο της.Με αυτό τον τρόπο της μηχανικής μάθησης το σύστημα δεν καθοδηγείται από κανέναν για το ποια ενέργεια θα πρέπει να ακολουθήσει αλλά προσπαθεί και πρέπει να ανακαλύψει μόνο του ποιες ενέργειες είναι αυτές που θα του αποφέρουν το μεγαλύτερο κέρδος.Ένα χαρακτηριστικό παράδειγμα για να γίνει κατανοητός ο συγκεκριμένος τρόπος είναι η εκμάθηση ενός συστήματος να παίζει ένα παιχνίδι εναντίον κάποιου αντιπάλου.

Χρησιμοποιείται σε προβλήματα: Σχεδιασμού (Planning), όπως για παράδειγμα ο έλεγχος κίνησης ρομπότ και η βελτιστοποίηση εργασιών σε εργοστασιακούς χώρους.

Για κάθε πρόβλημα προς επίλυση στο χώρο της Μηχανικής Μάθησης υπάρχει ένας κατάλληλος τρόπος μάθησης και για κάθε τρόπο μάθησης υπάρχει τουλάχιστον ένας κατάλληλος αλγόριθμος που μπορεί να χρησιμοποιηθεί [2].

Κεφάλαιο 3

Εισαγωγή στα νευρωνικά δίκτυα (Neural Network)

3.1 Ιστορική αναδρομή

Τα νευρωνικά δίκτυα άρχισαν να αναπτύσσονται ιδιαίτερα από τις αρχές τις δεκαετίας του 80.Το πρώτο τεχνητό νευρωνικό δίκτυο το δημιούργησε ο νευροφυσιολόγος, Warren McCulloch το 1943.Η περιορισμένη ανάπτυξη της τεχνολογίας όμως δε του επέτρεψε να το αναπτύξει παραπάνω. Οι πρώτες προσπάθειες έγιναν από τους McCulloch και Pitts οι οποίοι δημιούργησαν μοντέλα νευρωνικών δικτύων βασιζόμενοι στις γνώσεις τους στην νευρολογία.Τα δίκτυά τους βασίζονταν σε απλούς νευρώνες που θεωρούνταν ότι ήταν δυαδικές διατάξεις με σταθερά όρια.Τα αποτελέσματα των μοντέλων τους ήταν απλές λογικές συναρτήσεις όπως ¨α¨ ή ¨β¨ και ¨α¨ και ¨β¨.Επιπλέον έγινε και άλλη μία προσπάθεια από δύο ομάδες, τους Farley και Clark το 1954 και τους Rochester, Holland, Haibit και Duda το 1956. Αποκορύφωμα αποτέλεσε η εφεύρεση του μοντέλου του απλού αισθητήρα (Perceptron) από τον Frank Rosenblatt το 1958.Ο αισθητήρας είχε τρία στρώματα νευρώνων και μπορούσε να συνδέσει τη μία δοθείσα είσοδο με μία τυχαία μονάδα εξόδου.Το μοντέλο Perceptron μπορούσε μεταβάλλοντας τις τιμές των βαρών των συνδέσεων του, να εκπαιδευτεί στην ταξινόμηση, σε κατηγορίες συγκεκριμένων παραδειγμάτων. Το 1969 οι Minsky και Papert δημοσιεύουν το βιβλίο τους με τίτλο ¨The Perceptron¨, στο οποίο αποδεικνύουν μαθηματικά ότι τα τεχνητά νευρωνικά δίκτυα ενός επιπέδου(απλού αισθητήρα) παρουσιάζουν σημαντικούς περιορισμούς στην

επίλυση συγκεκριμένων ζητημάτων, με αποτέλεσμα για ένα αρκετά μεγάλο χρονικό διάστημα οι περιορισμοί και οι δυσκολίες που δημιουργήθηκαν, απέτρεψαν πληθώρα ερευνητών να μελετήσουν τα τεχνητά νευρωνικά δίκτυα. Το 1974 ο Paul Werbos ανέπτυξε και χρησιμοποίησε τη μέθοδο της οπίσθιας τροφοδότησης (backpropagation). Πέρασαν αρκετά χρόνια μέχρι να γίνει γνωστή αυτή η προσέγγιση. Σήμερα, τα δίκτυα αυτά αποτελούν την πιο γνωστή εφαρμογή των νευρωνικών δικτύων. Τα τέλη του 1970 και στις αρχές της δεκαετίας του 1980 ήταν σημαντική η πρόοδος που σημειώθηκε στον τομέα των νευρωνικών δικτύων. Τα μέσα ενημέρωσης βοήθησαν αρκετά στη διάδοση αυτής της νέας τεχνολογίας. Τέλος, πρέπει να σημειωθεί ότι τα νευρωνικά δίκτυα διδάσκονται πλέον στα περισσότερα πανεπιστήμια και η έρευνα προχωράει σε πολλά μέτωπα.[3] [4]

3.2 Ορισμός και Λειτουργία

Ένα νευρωνικό δίκτυο αποτελείται από απλούς υπολογιστικούς κόμβους (νευρώνες) που είναι διασυνδεδεμένοι μεταξύ τους. Η διασύνδεση των κόμβων μπορεί να παρομοιαστεί με το κεντρικό νευρικό σύστημα (ΚΝΣ). Στην περίπτωση των βιολογικών νευρώνων, είναι μέρος του νευρικού ιστού, ενώ στην περίπτωση των τεχνητών νευρώνων, είναι ένα αφηρημένο αλγοριθμικό δημιούργημα.

Οι νευρώνες είναι τα δομικά στοιχεία του δικτύου. Κάθε νευρώνας δέχεται ένα σύνολο αριθμητικών εισόδων από διαφορετικές πηγές, είτε από άλλους νευρώνες είτε από το περιβάλλον. Με βάση αυτές τις εισόδους, εκτελεί έναν υπολογισμό και παράγει μια έξοδο. Αυτή η έξοδος είτε τροφοδοτείται ως είσοδος σε άλλους νευρώνες του δικτύου είτε αποτελεί την τελική έξοδο [5].

Υπάρχουν τρεις τύποι νευρώνων:

1.**Οι νευρώνες εισόδου**:Οι οποίοι δεν επιτελούν κανέναν υπολογισμό και το μόνο που κάνουν είναι να συνδέουν τις περιβαλλοντικές εισόδους του δικτύου με τους υπολογιστικούς νευρώνες.

2.**Οι νευρώνες εξόδου**:Είναι υπεύθυνοι για να διοχετεύουν στο περιβάλλον τις τελικές αριθμητικές εξόδους του δικτύου.

3.**Οι υπολογιστικοί νευρώνες ή κρυμμένοι νευρώνες**:Οι

οποίοι καλούνται να πολλαπλασιάσουν κάθε είσοδό τους με το αντίστοιχο συναπτικό βάρος και να υπολογίσουν το ολικό άθροισμα των γινομένων.Το άθροισμα αυτό τροφοδοτείται ως όρισμα στη συνάρτηση ενεργοποίησης(activation function), την οποία υλοποιεί εσωτερικά κάθε κόμβος.Η συνάρτηση ενεργοποίησης κανονικοποιεί τις τιμές εξόδου, δηλαδή τις αντιστοιχεί σε μια πεπερασμένη τιμή στο διάστημα [0,1] ή [-1,1].Η τελική τιμή του νευρώνα είναι και η έξοδος της συνάρτησης ενεργοποίησης. Εάν x_{ki} είναι η i-οστή είσοδος του k νευρώνα, w_{ki} το i-οστό συναπτικό βάρος του k νευρώνα και $\varphi(\cdot)$ η συνάρτηση ενεργοποίησης του νευρωνικού δικτύου, τότε η έξοδος y_k του k νευρώνα δίνεται από την εξίσωση:

$$y_k = \varphi\left(\sum_{i=0}^{N} x_{ki} \cdot w_{ki}\right)$$

Στον k-οστό νευρώνα υπάρχει ένα συναπτικό βάρος w_{k0} το οποίο ονομάζεται πόλωση ή κατώφλι (bias,threshold).Η τιμή της εισόδου του είναι πάντα η μονάδα, $x_{k0} = 1$.Εάν το συνολικό άθροισμα από τις υπόλοιπες εισόδους του νευρώνα είναι μεγαλύτερο από την τιμή αυτή, τότε ο νευρώνας ενεργοποιείται. Εάν είναι μικρότερο, τότε παραμένει ανενεργός.

3.3 Μοντέλο Τεχνητού νευρώνα

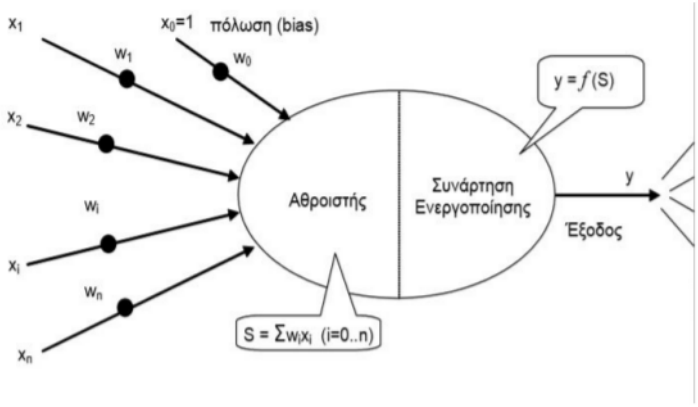

Figure 3.1: Μοντέλο Τεχνητού Νευρώνα
"Πηγή:NEURAL NETWORKS AND LEARNING MACHINES"

Ο τεχνητός νευρώνας αποτελεί δομικό συστατικό ενός νευρωνικού

δικτύου.Στο σχήμα 2.1 φαίνεται το μοντέλο ενός τεχνητού νευρώνα.Τα βασικά στοιχεία αυτού του μοντέλου είναι[6]:

1. Τα σήματα εισόδου χ_1, χ_2 , ..., χ_n που δέχεται και τα οποία είναι συνεχείς μεταβλητές.Κάθε σήμα εισόδου μεταβάλλεται από μία αρνητική ή θετική τιμή βάρους w_i,αντίστοιχο των συνάψεων.Οι συνάψεις αποτελούν τον τρόπο σύνδεσης των νευρώνων μεταξύ τους και χαρακτηρίζονται από το αντίστοιχο συναπτικό βάρος κάθε σήματος εισόδου.Συγκεκριμένα, ένα σήμα x_i στην είσοδο της σύναψης i που συνδέεται στον νευρώνα k, πολλαπλασιάζεται με το συναπτικό βάρος w_{ki} και έχει στόχο να ενισχύσει ή να αποδυναμώσει το συγκεκριμένο σήμα.

2. Η πόλωση(bias), η οποία είναι ειδική περίπτωση βάρους w_0 που επιδρά σε τιμή εισόδου $\chi_0 = 1$.Πρόκειται για ένα εξωτερικό σήμα εισόδου και χρησιμοποιείται συχνά για να διαμορφώσει κατάλληλα το κατώφλι της συνάρτησης ενεργοποίησης.

3. Ο αθροιστής, ο οποίος είναι υπεύθυνος για την πρόσθεση των επηρεασμένων από τα βάρη σημάτων και παράγει την ποσότητα S.Αυτές οι λειτουργίες αποτελούν το γραμμικό συνδυαστή.

4. Τη συνάρτηση ενεργοποίησης ή κατωφλίου (activation ή threshold function) η οποία δρα σαν φίλτρο, ορίζοντας ένα κατώφλι,το οποίο θα διαμορφώσει την τελική τιμή της εξόδου y,σε συνάρτηση με την ποσότητα S.

5. Την έξοδο η οποία μπορεί να αποτελεί είσοδο σε άλλους νευρώνες.

Έτσι, για κάθε νευρώνα k θα ισχύει:

$$y_k = \varphi\left(\sum_{i=0}^{N} x_{ki} \cdot w_{ki}\right)$$

3.4 Συναρτήσεις Ενεργοποίησης (Activations Functions)

Υπάρχουν διάφορες συναρτήσεις ενεργοποίησης (activation functions) ή συναρτήσεις μεταφοράς (transfer functions) f(.) που χρησιμοποιούνται στους τεχνητούς νευρώνες για να περιορίσουν την έξοδο τους εντός ενός κλειστού

μοναδιαίου διαστήματος [0,1] ή [-1,1]. [4].Οι περισσότεροι τύποι αυτών των συναρτήσεων χαρακτηρίζονται από μη γραμμικές συναρτήσεις ώστε να μπορούν να μοντελοποιούν μη γραμμικά φαινόμενα.Η συνάρτηση ενεργοποίησης μπορεί να πάρει διάφορες μορφές (γραμμική / μη γραμμική, παραγωγίσιμη στο πεδίο ορισμού της ή όχι) και η επιλογή γίνεται ανάλογα με τις επιθυμητές ιδιότητες του κάθε δικτύου.Διαφορετικές συναρτήσεις ενεργοποίησης εξάγουν διαφορετικές εξόδους. Οι 3 πιο βασικές συναρτήσεις ενεργοποίησης είναι οι παρακάτω [7]:

1.**Βηματική συνάρτηση (step transfer function) ή Συνάρτηση Κατωφλίου (Threshold Function)** :

Η συνάρτηση κατωφλίου χρησιμοποιείται κυρίως για να διαχωρίζει τις εξόδους του νευρώνα σε δύο επιμέρους κατηγορίες. Όταν οι έξοδοι αυτοί περιορίζονται στις δυαδικές τιμές 0 και 1, δηλαδή 0,1, τότε η συνάρτηση κατωφλίου αναφέρεται ως βηματική συνάρτηση (step function) και αντίστοιχα ως συνάρτηση προσήμου (signum function ή hardlimiter) όταν οι έξοδοι περιορίζονται στις τιμές -1 και 1, -1,1.

Η μαθηματική διατύπωση της βηματικής συνάρτησης είναι η εξής :

$$f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Η συνάρτηση f(x) είναι ασυνεχής εφόσον δεν ορίζεται στο x=0. Είναι η πιο απλή συνάρτηση ενεργοποίησης. Επειδή όμως δεν είναι παραγωγίσιμη συνάρτηση δεν χρησιμοποιείται αρκετά συχνά.

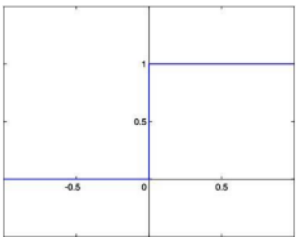

Figure 3.2: Βηματική Συνάρτηση (step function)
"Πηγή : Μπάστας Ν.,2007"

Αντίστοιχα η μαθηματική διατύπωση της συνάρτησης προσήμου είναι η εξής:

$$f(x) = \begin{cases} 1 & x > 0 \\ -1 & x \leq 0 \end{cases}$$

2.**Γραμμική συνάρτηση ενεργοποίησης (linear transfer function)** : Αυτού του τύπου η συνάρτηση χρησιμοποιείται κυρίως σε τεχνητούς νευρώνες που προορίζονται για γραμμική προσέγγιση στα γραμμικά φίλτρα.Στο επίπεδο εισόδου συνηθίζεται να χρησιμοποιείται μια γραμμική συνάρτηση μεταφοράς ενώ στο επίπεδο εξόδου και στο κρυφό επίπεδο χρησιμοποιείται μια σιγμοειδής συνάρτηση.Το μοντέλο εκπαιδεύεται πρώτα χρησιμοποιώντας δεδομένα εκπαίδευσης και στη συνέχεια επικυρώνεται χρησιμοποιώντας το τμήμα των δεδομένων που παρέμειναν [3]. Δίνεται από τον τύπο: f(x)=x

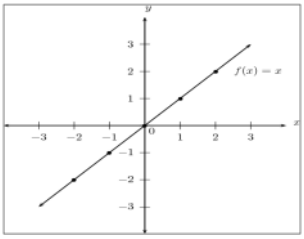

Figure 3.3: Γραμμική Συνάρτηση
"Πηγή : Σταθοπούλου Δ.,2010"

3.**Μη γραμμική συνάρτηση ενεργοποίησης (non-linear transfer function)** : Η μη γραμμική συνάρτηση ενεργοποίησης που χρησιμοποιείται συνήθως στα νευρωνικά δίκτυα ονομάζεται και σιγμοειδής συνάρτηση.Η σιγμοειδής συνάρτηση είναι εκείνη που χρησιμοποιείται περισσότερο στην κατασκευή των ΤΝΔ.Πρόκειται για μια αυστηρά αύξουσα συνάρτηση που διατηρεί ισορροπία ανάμεσα στην γραμμική και στην μη γραμμική συμπεριφορά.Οι τυπικές σιγμοειδείς είναι δύο:

Λογιστική σιγμοειδής:

$$f(x) = \frac{1}{1 + e^{-x}}$$

Υπερβολική εφαπτομένη: $f(x) = \tanh(x)$

Οι νευρώνες σε ένα τεχνητό νευρωνικό δίκτυο χρησιμοποιούν μη γραμμικές συναρτήσεις για να συμβάλουν στη δημιουργία της εξόδου του. Η

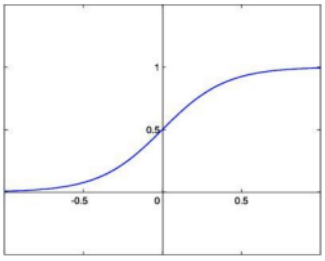

Figure 3.4: Λογιστική Σιγμοειδής
"Πηγή : Λιβιέρης Ι.,2008"

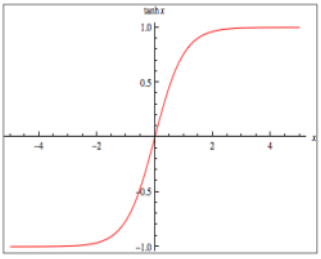

Figure 3.5: Υπερβολική Εφαπτομένη
"Πηγή : Λιβιέρης Ι.,2008"

σύνδεση μεταξύ των διαφόρων επιπέδων ενός δικτύου καθορίζεται από τα συναπτικά βάρη. Αυτά τα συγκεκριμένα βάρη που αναφέρονται έχουν ρυθμιστεί σύμφωνα με έναν αλγόριθμο μάθησης για τη σωστή και κατάλληλη εκπαίδευση ενός νευρωνικού δικτύου [8].

3.5 Αρχιτεκτονική Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ)

Ο όρος αρχιτεκτονική αναφέρεται στη δομή των νευρώνων που απαρτίζουν το δίκτυο και στις συνδέσεις μεταξύ τους.Η αρχιτεκτονική παίζει σημαντικό ρόλο στη λειτουργία και στην απόδοση ενός δικτύου.Όπως φαίνεται στην παραπάνω εικόνα, τα τεχνικά νευρωνικά δίκτυα είναι συνήθως οργανωμένα σε επίπεδα τα οποία ονομάζονται και στρώματα, και το κάθε επίπεδο (στρώμα) επεξεργάζεται ένα σύνολο σημάτων.Τα ενδιάμεσα επίπεδα ονομάζονται κρυμμένα επίπεδα και δεν είναι υποχρεωτικά.Τα επίπεδα αποτελούνται από έναν αριθμό μονάδων ή κόμβων που είναι συνδεδεμένες μεταξύ τους, έτσι ώστε μία μονάδα να έχει συνδέσμους με πολλές άλλες

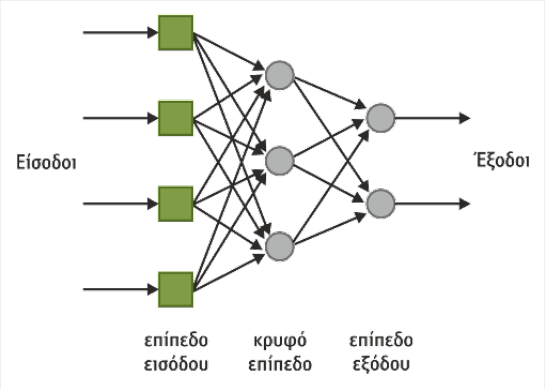

Figure 3.6: Αρχιτεκτονική Τεχνητών Νευρωνικών Δικτύων
"Πηγή: [2]"

μονάδες του ίδιου ή άλλου επιπέδου.Οι μονάδες επηρεάζουν άλλες μονάδες διεγείροντας ή αναστέλλοντας την ενεργοποίησή τους.Οι είσοδοι παρουσιάζονται στο δίκτυο μέσω του επιπέδου εισόδου το οποίο επικοινωνεί με έναν ή περισσότερα κρυφά επίπεδα.Τα κρυφά επίπεδα συνδέονται με το επίπεδο εξόδου από το οποίο εξάγεται η απάντηση.[9]

Υπάρχουν δύο βασικές κατηγορίες Τεχνικών Νευρωνικών Δικτύων σχετικά με τον τρόπο που συνδέονται οι μονάδες μεταξύ τους [7]:

• **Πρόσθιας τροφοδότησης(feed forward)**: Στα νευρωνικά δίκτυα πρόσθιας τροφοδότησης,οι μονάδες είναι οργανωμένες σε διαφορετικά επίπεδα έτσι ώστε οι μονάδες του ενός επιπέδου να τροφοδοτούν τις μονάδες του επόμενου επιπέδου,μέχρι να τροφοδοτηθούν και οι μονάδες του τελευταίου επιπέδου.Αυτό σημαίνει ότι δεν υπάρχει έξοδος μονάδας ενός επιπέδου που να αποτελεί είσοδο μονάδας του ίδιου ή προηγούμενων επιπέδου.Τέτοια τεχνικά νευρωνικά δίκτυα είναι γνωστά ως δίκτυα οπισθοδιάδοσης (backpropagation).

• **Οπίσθιας τροφοδότησης (feed backward)**: Στα οπισθίως τροφοδοτούμενα δίκτυα,που ονομάζονται και ανατροφοδοτούμενα ΤΝΔ (recurrent ANN),επιτρέπεται στις μονάδες ενός επιπέδου να τροφοδοτούν και μονάδες του ιδίου επιπέδου ή και προηγούμενων επιπέδων.Αν η ανατροφοδότηση αφορά κόμβους στο ίδιο επίπεδο,τότε τα δίκτυα ονομάζονται αυτοσυσχετιζόμενες μνήμες (autoassociated memories) διαφορετικά, ονομάζονται ετεροσυσχετιζόμενες μνήμες (heteroassociated memories).Στα ανατροφοδοτούμενα ΤΝΔ δεν υπάρχουν συνήθως παραπάνω από ένα κρυφό επίπεδο.Παρ' όλο που τα ανατροφοδοτούμενα δίκτυα είναι πολύ χρήσιμα,τα περισσότερα των νευρωνικών δικτύων είναι πρόσθιας τροφοδότησης.

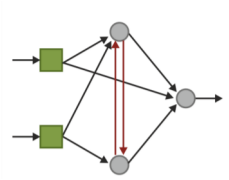

Figure 3.7: Αυτο-συσχετιζόμενη Μνήμη
"Πηγή: [2]"

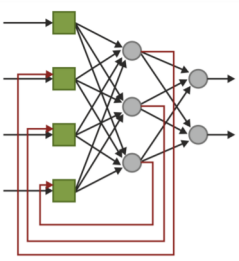

Figure 3.8: Ετερο-συσχετιζόμενη Μνήμη
"Πηγή: [2]"

Επιπλέον, οι νευρώνες σε ένα νευρωνικό δίκτυο, μπορούν να χαρακτηριστούν ως μερικώς ή πλήρως συνδεδεμένοι.Έτσι,εάν όλοι οι νευρώνες ενός επιπέδου συνδέονται με όλους τους υπόλοιπους,τότε χαρακτηρίζονται ως πλήρως συνδεδεμένοι (fully connected).Ενώ σε κάθε άλλη περίπτωση χαρακτηρίζονται ως μερικώς συνδεδεμένοι (partially connected).Παράδειγμα μερικής σύνδεσης αποτελούν τα δίκτυα με πρόσθια προώθηση (feedforward).Στα συγκεκριμένα,οι νευρώνες ενός επιπέδου συνδέονται πλήρως με τους νευρώνες στο επόμενο επίπεδο,ενώ δεν υπάρχουν συνδέσεις μεταξύ των νευρώνων ενός επιπέδου και του προηγουμένου στο νευρωνικό δίκτυο [8].

Τα νευρωνικά δίκτυα χωρίζονται στις παρακάτω κατηγορίες :

• **Δίκτυα Εμπρόσθιας Τροφοδότησης Ενός Επιπέδου (Single Layer Feedforward networks/ Single Layer Perceptron)**

•**Δίκτυα Εμπρόσθιας Τροφοδότησης Πολλών Επιπέδων (Multilayer Feedforward Networks/Multilayer Perceptrons MLP)**

• **Αναδρομικά δίκτυα (Recurrent Networks)**

• **Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural**

Networks)

3.6 Προβλήματα στην αναγνώριση αντικειμένων

Παρά το γεγονός ότι το μοντέλο του τεχνικού νευρωνικού δικτύου (Artificial Neural Network - ANN) έχει αποδειχθεί χρήσιμο σε πολλές εφαρμογές, δεν είναι τόσο αποτελεσματικό όταν πρόκειται για την αναγνώριση αντικειμένων σε εικόνες. Υπάρχουν διάφοροι λόγοι για αυτό, όπως[10]:

1. **Τοπολογία**: Ένα πλήρως συνδεδεμένο ANN αδυνατεί να λάβει υπόψη την τοπολογία της εισόδου. Μια εικόνα έχει τοπικά πολύ ισχυρή χωρική συσχέτιση μεταξύ των γειτονικών pixels, γεγονός που επιτρέπει τον συνδυασμό χαρακτηριστικών χαμηλής τάξης (ακμές, γωνίες κτλ.) μιας περιοχής σε χαρακτηριστικά υψηλής τάξης (π.χ. μύτες, αυτιά κ.ά.). [11]

2. **Κλιμάκωση**: Ακόμα και σχετικά μικρές εικόνες περιέχουν συνήθως μεγάλο αριθμό pixels (δηλ. εισόδων στο δίκτυο). Π.χ. μια εικόνα 32x32 περιέχει 1024 pixels. Ένα τυπικό, πλήρως συνδεδεμένο ANN με π.χ. 100 κρυφούς νευρώνες θα έπρεπε να υπολογίζει 1024x100 βάρη στο πρώτο επίπεδο. Αυτό καθιστά την κλιμάκωση σε μεγαλύτερες εικόνες δύσκολη και μη αποδοτική. [11]

3. **Μεταβλητότητα αντικειμένων**: Κάποια αντικείμενα μπορεί να είναι αρκετά όμοια σε ένα υψηλό επίπεδο, ώστε να μπορούν να ανήκουν στην ίδια κλάση, όμως ταυτόχρονα να είναι αρκετά διαφορετικά μεταξύ τους σε ένα χαμηλότερο επίπεδο. Για παράδειγμα, το ανθρώπινο πρόσωπο έχει αρκετά χαρακτηριστικά που ορίζουν ότι είναι πρόσωπο π.χ. μάτια, στόμα, μύτη κτλ.Όμως, το μέγεθος και το σχήμα αυτών των χαρακτηριστικών τείνουν να διαφέρουν πολύ από άτομο σε άτομο.Για να μπορέσει ένα τυπικό ANN να αντισταθμίσει αυτές τις εσωτερικές διαφορές εντός μιας κλάσης, θα πρέπει να κάνει τρεις κοστοβόρες προσαρμογές. α) Το δίκτυο θα πρέπει να είναι πολύ μεγάλο, β) πιθανότατα θα περιέχει νευρώνες με παρόμοια διανύσματα βαρών τοποθετημένους σε διαφορετικές θέσεις του δικτύου και γ) θα απαιτεί τεράστιο αριθμό δειγμάτων εκπαίδευσης.[11]

3.7 Κατηγορίες τεχνητών νευρωνικών δικτύων

3.7.1 Ο Αισθητήρας Perceptron

Ο αισθητήρας Perceptron είναι από τα παλαιότερα μοντέλα νευρωνικών δικτύων που αναπτύχθηκαν και είναι μία τοπολογία πρόσθιας τροφοδότησης, χωρίς κρυφά επίπεδα.Σήμερα υπάρχουν πολλές παραλλαγές με διαφορετικές νευρωνικές δομές,αλλά η πιο απλή είναι ο στοιχειώδης αισθητήρας (elementary perceptron) ο οποίος αποτελείται μόνο από ένα επίπεδο υπολογιστικών νευρώνων που ακολουθούν το μοντέλο Mc Culloch-Pitts.Χρησιμοποιεί ως συνάρτηση ενεργοποίησης τη βηματική συνάρτηση και χρησιμοποιείται για το διαχωρισμό ή την ταξινόμηση ενός συνόλου δεδομένων σε δύο κλάσεις με τη χρήση επιβλεπόμενης μάθησης. [12].

Σε περίπτωση που υπάρχει ένας αισθητήρας και δύο πιθανές κλάσεις, τα συναπτικά βάρη προσαρμόζονται συνεχώς μέχρι η διαφορά μεταξύ του επιθυμητού αποτελέσματος και του πραγματικού αποτελέσματος για κάθε περίπτωση να γίνει μηδέν. Αυτός ο κανόνας ονομάζεται "Κανόνας Δέλτα" και το σύμβολο δ αναφέρεται στη διαφορά μεταξύ του επιθυμητού και του πραγματικού αποτελέσματος εξόδου.Η διαδικασία τροποποίησης των συναπτικών βαρών υλοποιείται ως εξής [10]:

• Εάν η πραγματική έξοδος είναι 0 ενώ η επιθυμητή έξοδος είναι 1,τότε προστίθεται η τιμή κάθε εισόδου στο αντίστοιχο συναπτικό βάρος.

• Εάν η πραγματική έξοδος είναι 1 ενώ η επιθυμητή είναι 0, τότε αφαιρείται η τιμή κάθε εισόδου στο αντίστοιχο συναπτικό βάρος.

Παρακάτω παρουσιάζεται σχηματικά η επαναληπτική διαδικασία εκμάθησης του αισθητήρα.

Έχει ένα αριθμό εισόδων και μία μόνο έξοδο.Αυτό σημαίνει ότι η μονάδα αυτή δέχεται πολλές εισόδους s1, s2, s3... sn αλλά παράγει μία μόνο έξοδο είτε 1 είτε 0.Κάθε εισερχόμενο σήμα,συνδέεται με τον κεντρικό νευρώνα και το w (weight) είναι η επίδραση του εισερχόμενου σήματος με τον νευρώνα αυτό.Σημαντικό ρόλο παίζει το γινόμενο $s_i \cdot w_i$, όπου κάθε s_i πολλαπλασιάζεται με το βάρος w_i που έχει η σύνδεση i με αποτέλεσμα αυτό που παρουσιάζεται στο νευρώνα από κάθε εισερχόμενο σήμα είναι το γινόμενο $s_i \cdot w_i$.

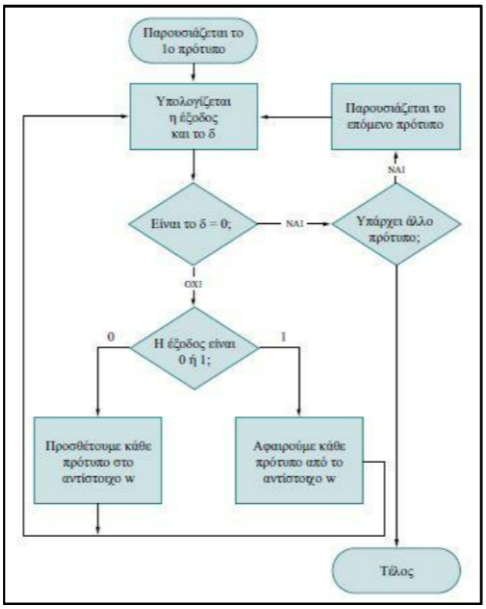

Figure 3.9: Σχηματική αναπαράσταση αλγορίθμου εκπαίδευσης perceptron "Πηγή: Αργυράκης Π. 2001"

Ο αισθητήρας στη συνέχεια αθροίζει αυτά τα γινόμενα και λαμβάνει ένα συνολικό σήμα με τιμή:

$S = \sum_i s_i \cdot w_i$

Τέλος, εφαρμόζουμε τη συνάρτηση κατωφλίου Heaviside με μια συγκεκριμένη τιμή κατωφλίου, η οποία συμβολίζεται ως ϑ.Συγκρίνουμε το ϑ με το άθροισμα S.Εάν S $>$ ϑ, τότε ο αισθητήρας ενεργοποιείται.Εάν S $<$ ϑ, τότε το άθροισμα S μηδενίζεται και ο αισθητήρας παραμένει ανενεργός.Αυτό μπορεί να συνοψιστεί ως εξής:

- Εάν S $>$ ϑ τότε η έξοδος ισούται με 1.
- Εάν S $<$ ϑ τότε η έξοδος ισούται με 0.

Καταλαβαίνουμε, ότι η δραστηριότητα του αισθητήρα εξαρτάται από τα βάρη των συνδέσεων,τις τιμές των εισόδων και την τιμή του κατωφλίου.

Ένα από τα πιο συνηθισμένα προβλήματα στα νευρωνικά δίκτυα είναι η συνάρτηση XOR.Αυτή η συνάρτηση δέχεται δύο εισόδους και παράγει μία έξοδο.Οι είσοδοι και η έξοδος μπορούν να είναι είτε 0 είτε 1 με τον ακόλουθο περιορισμό: Αν και οι δύο είσοδοι είναι ίδιες, τότε η έξοδος είναι 0, ενώ αν

είναι διαφορετικές, η έξοδος είναι 1. Αυτό το πρόβλημα δεν μπορεί να επιλυθεί με έναν μόνο νευρώνα, αλλά απαιτεί ένα δίκτυο νευρώνων.

Ένα στρώμα που αποτελείται από πολλές μονάδες perceptron ονομάζεται πυκνό στρώμα.

3.7.2 Δίκτυα Εμπρόσθιας Τροφοδότησης Ενός Επιπέδου

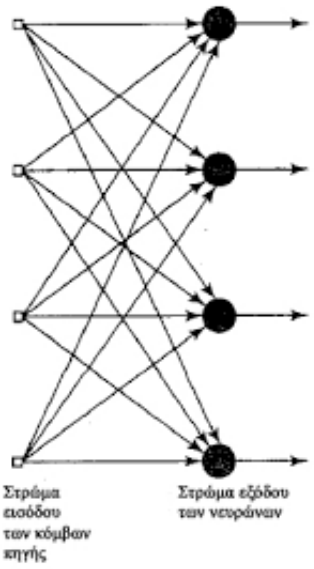

Figure 3.10: Δίκτυα Εμπρός Τροφοδότησης Ενός Επιπέδου
"Τροποποιήθηκε από την πηγή: ΧΑΤΖΗΧΡΙΣΤΑΚΗΣ Φ,2018"

Ένα νευρωνικό δίκτυο ενός επιπέδου είναι η απλούστερη μορφή νευρωνικού δικτύου.Στην περίπτωση αυτή, οι είσοδοι κάθε νευρώνα συνδέονται με τις αντίστοιχες εισόδους του δικτύου και η έξοδος κάθε νευρώνα είναι επίσης έξοδος του δικτύου, αλλά δεν ισχύει το αντίστροφο.Δεν είναι δυνατόν να πάμε από τους νευρώνες εξόδου στους κόμβους εισόδου.Σε αυτή την περίπτωση, το δίκτυο είναι ένα αυστηρά εμπρόσθιο δίκτυο και ονομάζεται εμπρόσθιο δίκτυο ενός επιπέδου.Με τον όρο "ένα επίπεδο", εννοούμε το επίπεδο εξόδου που περιέχει τους νευρώνες όπου εκτελούνται οι υπολογισμοί.Τέλος, πρέπει να σημειωθεί ότι το επίπεδο εισόδου με τους κόμβους εισόδου δεν υπολογίζεται, καθώς εκεί δεν εκτελούνται υπολογισμοί [11].

3.7.3 Δίκτυα Εμπρόσθιας Τροφοδότησης Πολλών Επιπέδων (MLP)

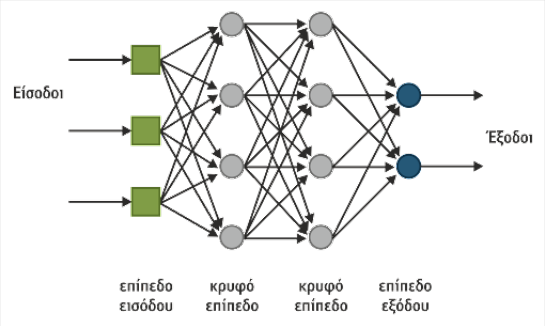

Figure 3.11: Δίκτυα Εμπρός Τροφοδότησης Πολλών Επιπέδων "Πηγή: [2]"

Το πολυστρωματικό perceptron αναφέρεται σε ένα νευρωνικό δίκτυο που συνδέει πολλά επίπεδα μεταξύ τους σε ένα κατευθυνόμενο γράφημα, δηλαδή η κίνηση του σήματος μέσω των κόμβων έχει μόνο μία κατεύθυνση.Το σήμα εισόδου περνάει από το ένα επίπεδο στο επόμενο και έτσι συνεχίζει προς τα μπροστά.Αυτό το δίκτυο εμπλουτίστηκε με την προσθήκη των κρυφών επιπέδων, τα οποία αποτελούνται τουλάχιστον από τρία επίπεδα: ένα επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα που αποτελούνται από νευρώνες που εκτελούν υπολογισμούς και ένα επίπεδο εξόδου που αποτελείται επίσης από νευρώνες που εκτελούν υπολογισμούς [13].

Κάθε κόμβος, εκτός από τους κόμβους εισόδου, διαθέτει μια μη γραμμική συνάρτηση ενεργοποίησης γιατί με την χρήση γραμμικών συναρτήσεων ενεργοποίησης είναι αδύνατος ο υπολογισμός πιο σύνθετων συνεχών και διαφορίσιμων συναρτήσεων.

Ο πιο συνηθισμένος τύπος δικτύων είναι τα δίκτυα τροφοδότησης προς τα εμπρός(feed-forward networks), στα οποία η πληροφορία ρέει από το επίπεδο εισόδου στο επίπεδο εξόδου χωρίς βρόχους.Τα MLP ανήκουν σε αυτή την κατηγορία δικτύων.Ένα δίκτυο τροφοδότησης προς τα εμπρός είναι ένα τεχνητό νευρωνικό δίκτυο όπου οι συνδέσεις μεταξύ των κόμβων δεν σχηματίζουν κύκλο, γεγονός που το καθιστά διαφορετικό από τα επαναλαμβανόμενα νευρωνικά δίκτυα.

Το νευρωνικό δίκτυο τροφοδοσίας ήταν ο πρώτος και απλούστερος τύπος τεχνητού νευρωνικού δικτύου που επινοήθηκε.Σε αυτό το δίκτυο, οι πληροφορίες μετακινούνται μόνο προς μία κατεύθυνση, προς τα εμπρός, από

τους κόμβους εισόδου, μέσω των κρυφών κόμβων (αν υπάρχουν) και στους κόμβους εξόδου.Δεν υπάρχουν κύκλοι ή βρόχοι στο δίκτυο.Ο στόχος του feedforward δικτύου είναι να βρεθούν τα κατάλληλα βάρη για την προσέγγιση μιας συνάρτησης.

Οι νευρώνες στα κρυφά στρώματα περιέχουν μια μη γραμμική συνάρτηση ενεργοποίησης που είναι συνεχής και παραγωγίσιμη σε κάθε σημείο. Επιπλέον, δεν υπάρχει σύνδεση μεταξύ νευρώνων του ίδιου στρώματος ή συνήθως μεταξύ νευρώνων που ανήκουν σε μη διαδοχικά στρώματα.Οι νευρώνες σε δίκτυα πολλαπλών στρωμάτων με τροφοδότηση προς τα εμπρός τροφοδοτούν με τα σήματα εξόδου τους τα επόμενα επίπεδα και τροφοδοτούνται από τα σήματα εξόδου των προηγουμένων επιπέδων.Στις περισσότερες περιπτώσεις, υπάρχει πλήρης διασύνδεση μεταξύ νευρώνων σε διαδοχικά στρώματα.

Κάθε κρυφός νευρώνας και κάθε νευρώνας εξόδου ενός MLP έχει σχεδιαστεί για να εκτελεί δύο υπολογισμούς:

1.Υπολογίζει το σήμα συνάρτησης που εμφανίζεται στην έξοδο του νευρώνα και εκφράζεται σαν μια συνεχής μη γραμμική συνάρτηση των σημάτων που εισέρχονται στο νευρώνα και των αντίστοιχων αρών των συνάψεων που σχετίζονται με το νευρώνα.

2.Υπολογίζει μια στιγμιαία προσέγγιση του διανύσματος κλίσης (δηλαδή τις κλίσεις της επιφάνειας σφάλματος ως προς τα άρη που σχετίζονται με τις συνάψεις που εισέρχονται στο νευρώνα), κατά την προς τα πίσω διάδοση του σήματος στο διαδίκτυο.

Στα δίκτυα perceptron πολλαπλών στρωμάτων, οι βηματικές συναρτήσεις δεν μπορούν να χρησιμοποιηθούν επειδή δεν είναι παραγωγήσιμες. Αυτό οφείλεται στο γεγονός ότι οι περισσότεροι κανόνες εκπαίδευσης βασίζονται σε μεθόδους βελτιστοποίησης που χρησιμοποιούν παραγώγους, οπότε αντί αυτών χρησιμοποιούνται συνήθως οι συναρτήσεις σιγμοειδούς και υπερβολικής εφαπτομένης. Το MLP είναι μια τεχνική βαθιάς μάθησης που χρησιμοποιεί πολλαπλά στρώματα νευρώνων και χρησιμοποιείται ευρέως για την επίλυση προβλημάτων μάθησης με επίβλεψη.Σχετικά παραδείγματα είναι οι εφαρμογές που περιλαμβάνουν την αναγνώριση ομιλίας, την αναγνώριση εικόνας και τη μηχανική μετάφραση.

Τα πολλαπλά επίπεδα και η μη γραμμική ενεργοποίησή του, διακρίνουν το MLP από ένα γραμμικό perceptron.

Το σχήμα 2.11 απεικονίζει την αρχιτεκτονική ενός MLP δικτύου με δύο κρυφά επίπεδα και πλήρη διασύνδεση μεταξύ τους.Πλήρως διασυνδεδεμένο ονομάζεται ένα δίκτυο του οποίου κάθε κόμβος είναι συνδεδεμένος με όλους τους κόμβους του προηγούμενου επιπέδου.Σε αυτό το είδος δικτύων, το σήμα μεταφέρεται σταδιακά από αριστερά προς τα δεξιά και από επίπεδο σε επίπεδο.

Στο δίκτυο μεταδίδονται δύο ειδών σήματα:

1. Σήματα συναρτήσεων (Function signals).Ένα σήμα συνάρτησης είναι ένα σήμα εισόδου (ερέθισμα) που ξεκινάει από τους κόμβους εισόδου του δικτύου, διαδίδεται προς τα μπροστά, από νευρώνα σε νευρώνα, και καταλήγει στους νευρώνες εξόδου του δικτύου.Σε κάθε νευρώνα του δικτύου από όπου περνάει το σήμα υπολογίζεται σαν συνάρτηση όλων των εισερχόμενων σημάτων και των αντίστοιχων αρών των συνάψεων που καταλήγουν στο συγκεκριμένο νευρώνα.

2. Σήματα σφάλματος (Error signals).Ένα σήμα σφάλματος, ξεκινάει από τους νευρώνες εξόδου του δικτύου και διαδίδεται προς τα πίσω από επίπεδο σε επίπεδο.Κάθε νευρώνας υπολογίζει το σήμα σφάλματος μέσω μιας συνάρτησης που εξαρτάται από το σφάλμα.

Αλγόριθμος Ανάστροφης Μετάδοσης Λάθους (Backpropagation)

Για την εκπαίδευση του MLP (Multi Layer Perceptron), στις περισσότερες περιπτώσεις χρησιμοποιείται η μέθοδος Backpropagation, η οποία βασίζεται στον κανόνα μάθησης διόρθωσης σφάλματος. Συγκεκριμένα, για κάθε είσοδο που δίνεται στο δίκτυο, οι έξοδοι κάθε μονάδας στο κρυφό στρώμα ή στο στρώμα εξόδου υπολογίζονται χρησιμοποιώντας συναρτήσεις μετάβασης. Για κάθε μονάδα εξόδου, οι διαφορές μεταξύ των υπολογισμένων και των επιθυμητών εξόδων λαμβάνονται υπόψη και διαδίδονται προς τα πίσω στις μονάδες των κρυφών στρωμάτων, προκειμένου να καθοριστούν οι απαραίτητες αλλαγές στα βάρη σύνδεσης μεταξύ των μονάδων. Με άλλα λόγια, αυτό μας επιτρέπει να προσδιορίσουμε τις απώλειες που υφίσταται κάθε κόμβος, ώστε να ενημερώσουμε τα βάρη έτσι ώστε να ελαχιστοποιηθεί η απώλεια, αναθέτοντας χαμηλότερα βάρη σε κόμβους με υψηλότερα ποσοστά σφάλματος και αντίστροφα. Αυτές οι αλλαγές αποσκοπούν στην όσο το δυνατόν μεγαλύτερη ελαχιστοποίηση των σφαλμάτων στην έξοδο.

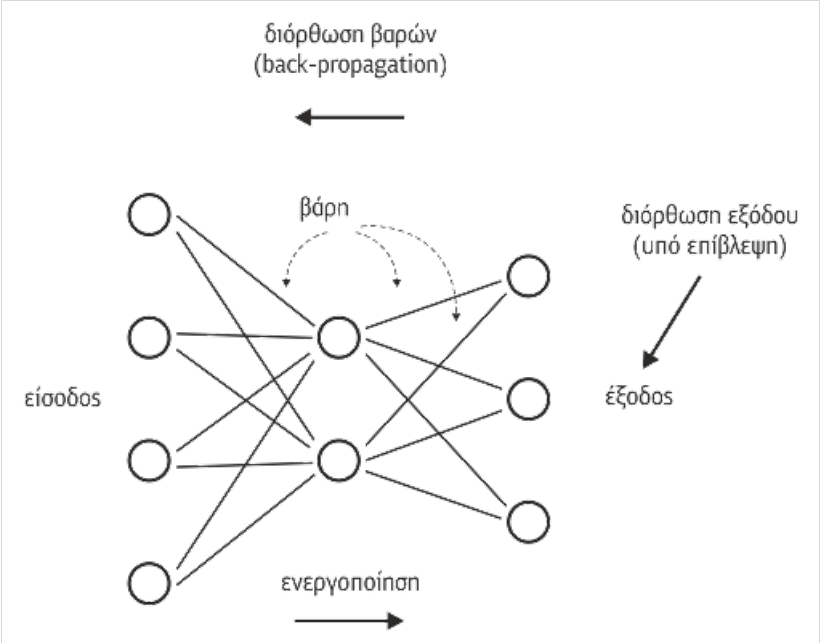

Figure 3.12: Εκπαίδευση με οπισθοδιάδοση
"Πηγή: [2]

3.7.4 Επαναλαμβανόμενο Νευρωνικό Δίκτυο (RNN)

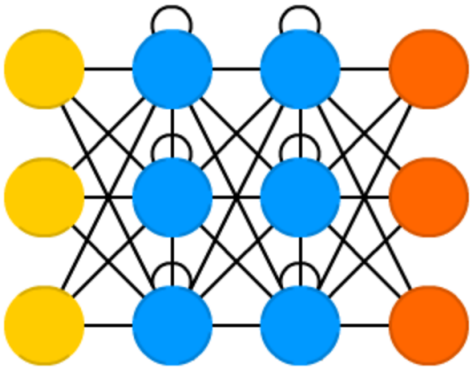

Figure 3.13: Επαναλαμβανόμενο Νευρωνικό Δίκτυο
"Πηγή: [14]"

Ένα επαναλαμβανόμενο νευρωνικό δίκτυο (RNN) είναι μια κατηγορία τεχνητών νευρωνικών δικτύων όπου οι συνδέσεις μεταξύ των μονάδων σχηματίζουν έναν κατευθυνόμενο κύκλο.Τα RNN ονομάζονται

επαναλαμβανόμενα επειδή εκτελούν την ίδια εργασία για κάθε στοιχείο μιας ακολουθίας, με την έξοδο να εξαρτάται από τους προηγούμενους υπολογισμούς.Το κύριο χαρακτηριστικό των RNN δικτύων είναι ότι διαθέτουν μνήμη την οποία εκμεταλλεύονται για να επεξεργάζονται αυθαίρετες ακολουθίες εισόδων.Όπως αναφέρθηκε και παραπάνω, τα ΤΝΔ έχουν κρυμμένα επίπεδα, έτσι λοιπόν και τα RNN χρησιμοποιούν αυτά τα επίπεδα για να έχουν αυτόνομες συνδέσεις και ενεργοποιούνται από το κατώτερο στρώμα και την προηγούμενη τιμή ενεργοποίησης. [15] [16]

Ο όρος επαναλαμβανόμενο νευρωνικό δίκτυο χρησιμοποιείται για να αναφερθεί σε δύο ευρείες κατηγορίες δικτύων.Η πρώτη είναι πεπερασμένης ώθησης και η δεύτερη άπειρης ώθησης.Και οι δύο κατηγορίες δικτύων παρουσιάζουν διαχρονική δυναμική συμπεριφορά.Ένα επαναλαμβανόμενο δίκτυο πεπερασμένης ώθησης είναι ένας κατευθυνόμενος άκυκλος γράφος που μπορεί να ξετυλιχθεί και να αντικατασταθεί με ένα αυστηρά εμπρόσθιο νευρωνικό δίκτυο, ενώ ένα επαναλαμβανόμενο δίκτυο άπειρης ώθησης είναι ένας κατευθυνόμενος κυκλικός γράφος που δεν μπορεί να ξετυλιχθεί.Τόσο τα επαναλαμβανόμενα δίκτυα πεπερασμένης ώθησης όσο και τα δίκτυα άπειρης ώθησης μπορούν να έχουν πρόσθετες αποθηκευμένες καταστάσεις και η αποθήκευση μπορεί να τεθεί υπό άμεσο έλεγχο από το νευρωνικό δίκτυο.Η αποθήκευση μπορεί επίσης να αντικατασταθεί από άλλο δίκτυο ή γράφημα, εάν αυτό ενσωματώνει χρονικές καθυστερήσεις ή έχει βρόχους ανατροφοδότησης.Τέτοιες ελεγχόμενες καταστάσεις αναφέρονται ως περιφραγμένη κατάσταση ή πύλη μνήμης και αποτελούν μέρος μακροχρόνιων δικτύων βραχυπρόθεσμης μνήμης (LSTM) και των επαναλαμβανόμενων μονάδων με πύλη.Αυτό ονομάζεται επίσης νευρωνικό δίκτυο ανατροφοδότησης (FNN).

Ένα δίκτυο μπορεί να είναι πλήρως συνδεδεμένο ή μερικώς συνδεδεμένο. Σε ένα πλήρως συνδεδεμένο δίκτυο οι νευρώνες του κάθε επιπέδου του συνδέονται με όλους τους νευρώνες του επόμενου επιπέδου, ενώ σε ένα μερικώς συνδεδεμένο δίκτυο συνδέονται με ορισμένους νευρώνες του επόμενου επιπέδου.

Η ιδέα πίσω από τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) είναι η δημιουργία ενός δικτύου που μπορεί να χρησιμοποιεί ακολουθίες ως εισόδους.Αυτές οι ακολουθίες μπορεί να έχουν διαφορετικά μήκη και οι τιμές τους μπορεί να σχετίζονται μεταξύ τους.[17] [14]

Η κύρια διαφορά των επαναλαμβανόμενων νευρωνικών δικτύων (RNN) από τα δίκτυα τροφοδότησης (MLP) είναι ότι περιέχουν βρόχους.Αυτή η

διαφορά διευκολύνει την εκπαίδευση των νευρώνων για τον χειρισμό μακροχρόνιων εξαρτήσεων στις ακολουθίες εισόδου, επιτρέποντάς τους να χρησιμοποιούν περαιτέρω πληροφορίες.Με αυτόν τον τρόπο, οι νευρώνες διαθέτουν μια εσωτερική μνήμη που δημιουργείται από την είσοδό τους, η οποία εξαρτάται από τις προηγούμενες τιμές.Σε μια πολύ μεγάλη ακολουθία, οι τιμές της μπορεί να έχουν μακροχρόνιες εξαρτήσεις.Σε αυτή την περίπτωση το επαναλαμβανόμενο νευρωνικό δίκτυο πρέπει να απομνημονεύσει τις ακολουθίες αλλά αν ο αλγόριθμος εκπαίδευσης δεν το επιτρέπει θα είναι δύσκολο να εκπαιδευτεί από αυτές.Μια λύση σε αυτό το πρόβλημα είναι η δημιουργία παραλλαγών των συνηθισμένων επαναλαμβανόμενων νευρωνικών δικτύων που ελέγχουν τον τρόπο υπολογισμού αυτών των κλίσεων.

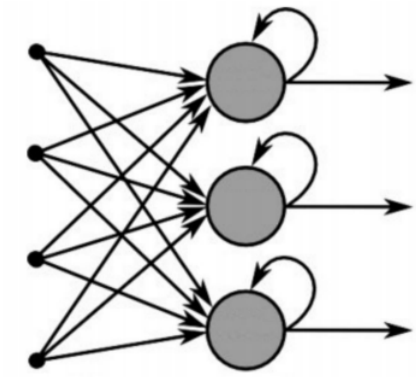

Figure 3.14: Επαναλαμβανόμενο Νευρωνικό Δίκτυο
"Τροποποιήθηκε από την πηγή: ΧΑΤΖΗΧΡΙΣΤΑΚΗΣ Φ,2018"

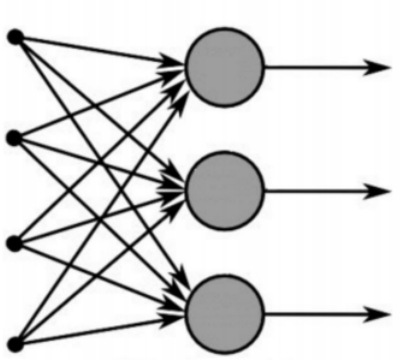

Figure 3.15: Δίκτυο Με Πρόσθια Τροφοδότηση
"Τροποποιήθηκε από την πηγή: ΧΑΤΖΗΧΡΙΣΤΑΚΗΣ Φ,2018"

3.7.5 Συνελικτικά Νευρωνικά Δίκτυα (CNN)

Ένα Συνελικτικό Νευρωνικό Δίκτυο (Convolutional Neural Network - CNN) είναι μια επέκταση του μοντέλου του τεχνητού νευρωνικού Δικτύου και έχει σχεδιαστεί ειδικά για την αναγνώριση αντικειμένων σε εικόνες ή για την αναγνώριση ομιλίας. Δημιουργήθηκε με σκοπό να λύσει τα θέματα που αντιμετώπιζε ένα τυπικό ANN μοντέλο(Artificial Neural Network) [11] [13].

Ο σκοπός των συνελικτικών δικτύων (Convolutional Neural Networks) είναι να "μάθουν" χαρακτηριστικά από τα δεδομένα μέσω συνελίξεων.Είναι ικανά να αναγνωρίζουν πρόσωπα, αντικείμενα, τοπία και πολλά άλλα είδη πραγμάτων.Επίσης, έχουν την δυνατότητα να αναγνωρίζουν ακόμη και ήχο.Τα δίκτυα αυτά αποτελούνται από πολλά επίπεδα τα οποία διαφέρουν μεταξύ τους ως προς την δομή και την λειτουργία που παρέχουν.

Το δίκτυο χωρίζεται σε τρία μέρη:

1. **Το επίπεδο εισόδου**

2. **Τα επίπεδα εξόρυξης χαρακτηριστικών** και

3. **Τα επίπεδα κατηγοριοποίησης/εξόδου**.

Η αποδεκτή είσοδος είναι τρισδιάστατη, δηλαδή μια εικόνα με τα κανάλια της και τα επίπεδα εξόρυξης χαρακτηριστικών έχουν συνήθως ένα επίπεδο συνέλιξης και στην συνέχεια ένα pooling επίπεδο.

Αυτά τα επίπεδα βρίσκουν τον αριθμό από χαρακτηριστικά που εμπεριέχονται στις εικόνες και σταδιακά εντοπίζουν πιο πολύπλοκα αντικείμενα.Τα χαρακτηριστικά εντοπίζονται με μικρά φίλτρα (συνήθως 3x3) τα οποία συνελλίσονται με την είσοδο.Κάθε φίλτρο είναι διαφορετικό και εντοπίζει διαφορετικά αντικείμενα.Τα αρχικά χαρακτηριστικά, είναι ακμές και γραμμές και σταδιακά αλλάζουν και φτάνουν στο σημείο να είναι περίπλοκα αντικείμενα τα οποία βοηθούν στην σωστή κατηγοριοποίηση.

Τα συνελικτικά δίκτυα αποτελούνται από πολλά επίπεδα τα οποία διαφέρουν μεταξύ τους ως προς την δομή και την λειτουργία που παρέχουν.Οι πιο συχνές κατηγορίες επιπέδων συνελικτικών δικτύων είναι:

- **Επίπεδο Συνέλιξης**:Η διαδικασία του επιπέδου συνέλιξης αποτελείται από μια σειρά φίλτρων τα οποία αναζητούν συγκεκριμένα χαρακτηριστικά ή μοτίβα στην αρχική εικόνα που εισέρχεται στο σύστημα. Τα

φίλτρα συνήθως έχουν μικρότερες διαστάσεις από την αρχική εικόνα, αλλά η διάσταση του βάθους παραμένει ίδια, δηλαδή 3. Κάθε φίλτρο κινείται κατά μήκος και πλάτος της εικόνας εισόδου, και υπολογίζεται ένα εσωτερικό γινόμενο για την παραγωγή ενός χάρτη ενεργοποίησης.Διαφορετικά φίλτρα τα οποία εντοπίζουν διαφορετικά χαρακτηριστικά περιστρέφονται στην εικόνα εισόδου και ένα σύνολο από χάρτες ενεργοποίησης προκύπτει ως έξοδος.Η σχέση που χρησιμοποιείται για να υπολογιστούν οι διαστάσεις των χαρτών ενεργοποίησης είναι η εξής [18]:

$(N + 2 * P - F)/S + 1$

Όπου:

N = διάσταση της εικόνας εισόδου

P = γέμισμα περιθωρίου με μηδενικά

F = διαστάσεις του φίλτρου

S = Άλμα

• **Pooling Επίπεδο**:Ο λόγος ύπαρξης αυτού του επιπέδου είναι η σταδιακή μείωση του μεγέθους των διαστάσεων καθώς η είσοδος είναι μια μεγάλη εικόνα ενώ η έξοδος αποτελείται από λίγες μόνο κλάσεις.Το συγκεκριμένο στρώμα χρησιμοποιεί την συνάρτηση max() για να αλλάξει το μέγεθος της εισόδου του.Αυτό γίνεται με τη χρήση ενός φίλτρου συνήθως μεγέθους 2x2, το οποίο εντοπίζει το μέγιστο των τεσσάρων αριθμών στους οποίους εφαρμόζεται. [19].

• **Πλήρως Συνδεδεμένο Επίπεδο**:Σε αυτό το επίπεδο όλοι οι νευρώνες συνδέονται πλήρως με όλες τις εξόδους των νευρώνων του προηγούμενου επιπέδου.Οι έξοδοι των συνελικτικών επιπέδων και των συγκεντρωτικών επιπέδων αντιπροσωπεύουν χαρακτηριστικά υψηλού επιπέδου από την αρχική εικόνα εισόδου.Ο σκοπός του συγκεκριμένου επιπέδου είναι να χρησιμοποιήσει αυτά τα χαρακτηριστικά για να ταξινομήσει την εικόνα εισόδου σε μία από τις διάφορες κλάσεις που υπάρχουν ως επιλογές.Εκτός από το ζήτημα της ταξινόμησης όμως, η χρήση αυτού του επιπέδου είναι συνήθως ένας τρόπος εκμάθησης των μη γραμμικών συνδυασμών αυτών των χαρακτηριστικών.Πολλά από τα χαρακτηριστικά που προκύπτουν από τα προηγούμενα στρώματα μπορεί να περιέχουν ήδη χρήσιμες πληροφορίες για το έργο της ταξινόμησης, αλλά ο συνδυασμός τους μπορεί να βελτιώσει περαιτέρω την απόδοση [19].

• **Softmax Επίπεδο**:Το επίπεδο softmax είναι το τελευταίο επίπεδο που υπάρχει στα συνελικτικά δίκτυα και είναι αυτό που κάνει την κατηγοριοποίηση.Βρίσκεται μετά το τελευταίο πλήρως διασυνδεδεμένο επίπεδο.Μετατρέπει τα διάφορα αποτελέσματα σε ένα διάνυσμα πιθανοτήτων έτσι ώστε το άθροισμα των επιμέρους στοιχείων να είναι μονάδα.

Οι συνελικτικοί νευρώνες τοποθετούνται συνήθως στα κατώτερα επίπεδα του δικτύου, όπου πραγματοποιείται η επεξεργασία των ανεπεξέργαστων εικονοστοιχείων εισόδου.Οι βασικοί νευρώνες βρίσκονται στα ανώτερα επίπεδα, όπου πραγματοποιείται η επεξεργασία της εξόδου από τα κατώτερα στρώματα.Τα κάτω στρώματα μπορούν συνήθως να εκπαιδευτούν μέσω μη επιβλεπόμενης μάθησης, χωρίς να στοχεύουν σε κάποια συγκεκριμένη εργασία πρόβλεψης.Τα βάρη τους θα εκπαιδευτούν να εντοπίζουν χαρακτηριστικά που εμφανίζονται συχνά στα δεδομένα εισόδου.Στην περίπτωση φωτογραφιών ζώων, τα τυπικά χαρακτηριστικά είναι τα αυτιά και οι μύτες, ενώ στην περίπτωση φωτογραφιών κτιρίων, τα χαρακτηριστικά είναι αρχιτεκτονικά στοιχεία όπως τοίχοι, στέγες και παράθυρα κ.ο.κ.Σε περίπτωση που ως δεδομένα εισόδου χρησιμοποιείται συνδυασμός διαφόρων αντικειμένων και σκηνών, τότε τα χαρακτηριστικά που θα μαθαίνονται στα κάτω επίπεδα θα είναι γενικού χαρακτήρα.Αυτό σημαίνει ότι τα προ-εκπαιδευμένα συνελικτικά επίπεδα μπορούν να χρησιμοποιηθούν σε πολλές διαφορετικές εργασίες επεξεργασίας εικόνας.Τα ανώτερα επίπεδα εκπαιδεύονται πάντα με τη χρήση τεχνικών μηχανικής μάθησης με επίβλεψη, όπως η οπισθοδιάδοση [20].

Συνολικά, ένα Συνελικτικό στρώμα:

1. Δέχεται ως είσοδο έναν όγκο μεγέθους $W1 \times H1 \times D1$.

2. Περιγράφεται από 4 παραμέτρους:

• Το μέγεθος των φίλτρων F.

• Τον αριθμό των φίλτρων K.

• Το βήμα του φιλτραρίσματος (stride) S.

• Τον αριθμό των μηδενικών που προστίθενται στα άκρα (padding) P.

3. Παράγει στην έξοδο του έναν όγκο μεγέθους $W2 \times H2 \times D2$ που θα αναφέρεται στο εξής ως όγκος χαρακτηριστικών ή χάρτης χαρακτηριστικών, όπου:

- W2 = $\lfloor \frac{W_1 - F + 2 \cdot P}{S} \rfloor + 1$
- H2 = $\lfloor \frac{H_1 - F + 2 \cdot P}{S} \rfloor + 1$
- D2 = K

4. Για κάθε φίλτρο χρειάζεται $F \cdot F \cdot D_1$ βάρη.Ο συνολικός αριθμός των παραμέτρων είναι $K \cdot (F \cdot F \cdot D_1) + K$, όπου ο τελευταίος όρος αντιστοιχεί στην πόλωση του κάθε φίλτρου.

3.8 Δίκτυα Μακράς Βραχύχρονης Μνήμης(LSTM)

Τα δίκτυα μακράς Βραχύχρονης μνήμης είναι μορφή τεχνητών ανατροφοδοτούμενων νευρωνικών δικτύων που χρησιμοποιούνται στον τομέα της βαθειάς μάθησης.Τα δίκτυα LSTM είναι κατάλληλα για την ταξινόμηση, την επεξεργασία και την πραγματοποίηση προβλέψεων με βάση δεδομένα χρονοσειρών, δεδομένου ότι μπορεί να υπάρχουν καθυστερήσεις άγνωστης διάρκειας μεταξύ σημαντικών γεγονότων σε μια χρονοσειρά (όπως δεδομένα ήχου ή βίντεο).Τα δίκτυα μακράς βραχύχρονης μνήμης χρησιμοποιούνται σε εφαρμογές όπως την αυτόματη ανάγνωση χειρογράφων, την αναγνώριση ομιλίας και την ανίχνευση ανωμαλιών σε δικτυακές επικοινωνίες.Τα LSTM έχουν σχεδιαστεί ειδικά για να αποφευχθεί το πρόβλημα της μακροχρόνιας εξάρτησης.Η ικανότητα "ανάμνησης" πληροφοριών για μεγάλες χρονικές περιόδους είναι η προεπιλεγμένη συμπεριφορά τους.Η βασική της μονάδα αποτελείται από ένα κελί, και τις πύλες εισόδου, εξόδου και λήθης.Το κελί κρατάει πληροφορίες από προηγούμενες θέσεις χρόνου ενώ οι τρεις πύλες ρυθμίζουν την ροή πληροφορίας εντός και εκτός του κελιού.Πιο συγκεκριμένα, η πύλη εισόδου αποφασίζει πόσες πληροφορίες θα διατηρηθούν στη μνήμη από το τελευταίο δείγμα, η πύλη εξόδου ρυθμίζει την ποσότητα των δεδομένων που περνούν στο επόμενο επίπεδο και οι πύλες λήθης ελέγχουν το ρυθμό διαγραφής της μνήμης που αποθηκεύεται [21].

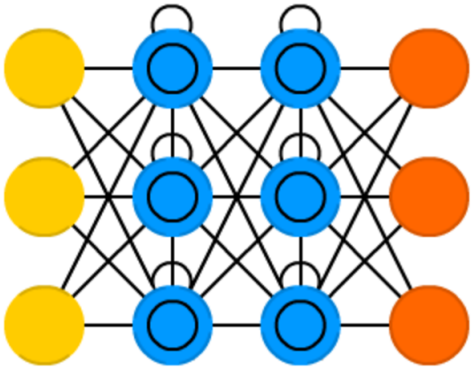

Figure 3.16: Δίκτυο Μακράς Βραχύχρονης Μνήμης
"Πηγή: [14]"

3.9 Εφαρμογές των τεχνητών νευρωνικών δικτύων

Τα τελευταία χρόνια, το ενδιαφέρον των ανθρώπων για τα νευρωνικά δίκτυα έχει αυξηθεί, γεγονός που είναι φυσικό, δεδομένου ότι εφαρμόζονται σε ένα ευρύ φάσμα επιστημονικών και τεχνολογικών τομέων, όπως τα οικονομικά, η ιατρική, η μηχανική, η γεωλογία, η φυσική, η ρομποτική κ.λπ. Στην πραγματικότητα, τα νευρωνικά δίκτυα εισάγονται οπουδήποτε απαιτείται πρόβλεψη, ταξινόμηση ή έλεγχος [2].

Οι βασικές λειτουργίες που μπορεί να εκτελέσει ένα τεχνητό νευρωνικό δίκτυο είναι οι παρακάτω:

1.Προσέγγιση συναρτήσεων: Είναι η βασική λειτουργία πάνω στην οποία βασίζεται και η δυνατότητα των νευρωνικών δικτύων για πρόβλεψη χρονοσειρών.Μια παραδειγματική περίπτωση θα μπορούσε να είναι η πρόβλεψη της κίνησης των επιβατών σε μια αεροπορική εταιρεία [22].

2. Αναγνώριση προτύπων: Με βάση τα δεδομένα που παρέχονται για ένα πρότυπο, το νευρωνικό δίκτυο μπορεί να αναγνωρίσει και να κατηγοριοποιήσει το πρότυπο αυτό.Η αναγνώριση προτύπων χρησιμοποιείται κυρίως σε διάφορους τομείς, όπως οι τράπεζες για την αυθεντικότητα υπογραφών, η πληροφορική και οι τηλεπικοινωνίες για την αναγνώριση ήχου και εικόνας.

3.Διαχωρισμός δεδομένων: Σε αυτήν τη διαδικασία, το δίκτυο

αναλύει σύνολα δεδομένων που του παρέχονται βάσει μιας κοινής ομοιότητας. Με αυτό τον τρόπο, το δίκτυο εκπαιδεύεται με ένα σωστό σετ δεδομένων, ώστε να μπορεί να διαχωρίσει οποιοδήποτε σετ δεδομένων του δοθεί σε ομάδες μέσω της διαδικασίας εκμάθησης.

Αντιπροσωπευτικά παραδείγματα προβλημάτων στα οποία η ανάλυση των νευρωνικών δικτύων έχει εφαρμοστεί με επιτυχία είναι τα εξής:

1.Ιατρική: Ιατρικές διαγνώσεις, ανάλυση και αναγνώριση καρκινικών όγκων, ανάπτυξη νέων φαρμάκων για ασθένειες.

2.Βιομηχανία: Πρόβλεψη πωλήσεων, πρόβλεψη παραγωγής, πρόβλεψη βιομηχανικών διεργασιών.

3.Οικονομία: Χρηματιστηριακές προβλέψεις, εξακρίβωση πιστοληπτικής ικανότητας.

4.Τεχνολογία: Δημιουργία ρομπότ για να διευκολύνεται η ζωή του ανθρώπου, δημιουργία μη επανδρωμένων αεροσκαφών και γενικότερα οχημάτων.

Κεφάλαιο 4

Ενισχυτική Μάθηση (Reinforcement Learning)

4.1 Εισαγωγή και Βασικά Στοιχεία

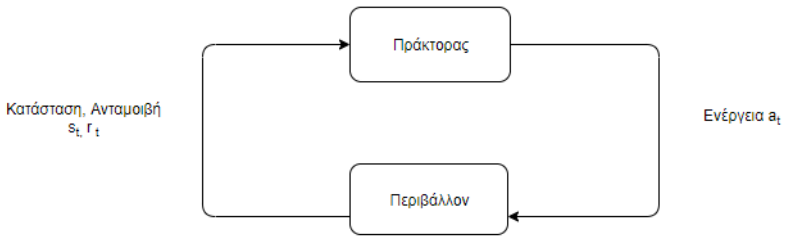

Figure 4.1: Q-learning
"Τροποποιήθηκε από την πηγή: [23]"

Η Ενισχυτική Μάθηση είναι η επιστήμη της λήψης βέλτιστων αποφάσεων χρησιμοποιώντας εμπειρίες.Η διαδικασία της ενισχυτικής μάθησης περιλαμβάνει τα εξής απλά βήματα [24] [25]:

1.Παρατήρηση του περιβάλλοντος.

2.Απόφαση για τον τρόπο δράσης με τη χρήση κάποιας στρατηγικής.

3.Ενεργεί αναλόγως.

4.Λήψη ανταμοιβής ή ποινής.

5.Μάθηση από τις εμπειρίες και βελτίωση της στρατηγικής μας επανάληψης μέχρι να βρεθεί η βέλτιστη στρατηγική.

Είναι μία τεχνική μηχανικής μάθησης όπου το σύστημα προσπαθεί να μάθει μέσα από την άμεση αλληλεπίδραση με το περιβάλλον.Πιο συγκεκριμένα, ο πράκτορας (agent) είναι μία οντότητα που συλλέγει πληροφορίες από το περιβάλλον του (environment) με απώτερο σκοπό να κατανοήσει την κατάσταση του περιβάλλοντός του και να λάβει το μέγιστο κέρδος ανταμοιβής σε σχέση με τις ενέργειες που πραγματοποιεί μέσα στο περιβάλλον.Η λήψη της ανταμοιβής είναι η μοναδική απάντηση ή αλλιώς το μοναδικό ερέθισμα που λαμβάνει ένας πράκτορας από την σύνδεση του με το περιβάλλον.Στη συνέχεια ο πράκτορας, ανάλογα με την τρέχουσα κατάσταση, ενεργεί προσπαθώντας να επιτύχει τους στόχους του.Το κατά πόσον ο πράκτορας πετυχαίνει τους στόχους του εκφράζεται από ένα αριθμητικό σήμα επιβράβευσης (reward signal).Επομένως, ο πράκτορας δρα με τέτοιο τρόπο ώστε να μεγιστοποιήσει το άθροισμα των σημάτων επιβράβευσης κατά τη διάρκεια της λειτουργίας του.Το σήμα αυτό είτε υπολογίζεται από τον ίδιο τον πράκτορα είτε δίνεται σε αυτόν από κάποια εξωτερική πηγή [26] [27].

Πιο αναλυτικά οι μέθοδοι για την ενισχυτική μάθηση αποτελούνται συνήθως από τα ακόλουθα στοιχεία [23] [28]:

S : Ένα σύνολο καταστάσεων (States).

A : Ένα σύνολο ενεργειών (Actions).

T : $S \times A$ -> Π(S) Μία συνάρτηση μετάβασης(transition function) που υπολογίζει την πιθανότητα μετάβασης από μια κατάσταση (state) πραγματοποιώντας μια ενέργεια (action) σε μια νέα κατάσταση (state).

R : $S \times A$ -> R Μία συνάρτηση επιβράβευσης (reward function) που υπολογίζει το reward για κάθε συνδυασμό state, action.

π : S -> A Μια πολιτική η οποία είναι μια συνάρτηση που υποδεικνύει ποια ενέργεια πρέπει να γίνει σε κάθε κατάσταση.

Ο κύριος στόχος της ενισχυτικής μάθησης είναι να υπολογιστεί η βέλτιστη πολιτική π^* η οποία και μεγιστοποιεί τα συνολικά rewards κατά την λειτουργία του πράκτορα.

Τα βασικά στοιχεία κάθε συστήματος ενισχυτικής μάθησης περιγράφονται παρακάτω:

• **Ο Πράκτορας**, ο οποίος συλλέγει πληροφορίες σχετικά με το περιβάλλον του και εκτελεί ενέργειες σύμφωνα με αυτές τις πληροφορίες. Σκοπός του πράκτορα είναι να μεγιστοποιήσει ένα σήμα ενίσχυσης (ανταμοιβή) το οποίο λαμβάνεται από το περιβάλλον ως αποτέλεσμα των ενεργειών του. Ο πράκτορας δεν λαμβάνει καμία άλλη ανατροφοδότηση από το περιβάλλον, εκτός από τις ανταμοιβές, πράγμα που σημαίνει ότι πρέπει να μάθει πως να συμπεριφέρεται με περιορισμένες πληροφορίες.

• **Το Περιβάλλον**, μέσα στο οποίο δρα ο πράκτορας. Περιβάλλον θεωρείται οτιδήποτε δεν είναι ο ίδιος ο πράκτορας. Το περιβάλλον, αναπαρίσταται σαν μια διαδικασία απόφασης Markov (Markov Decision Process) και για την επίλυση της χρησιμοποιούνται συχνά τεχνικές δυναμικού προγραμματισμού. Η ενισχυτική μάθηση στοχεύει σε σχετικά μεγάλες MDPs για τις οποίες η εφαρμογή των κλασσικών αλγορίθμων δυναμικού προγραμματισμού θα ήταν ανέφικτη.

• **Η πολιτική**, η οποία καθορίζει τον τρόπο συμπεριφοράς του πράκτορα σε μία δεδομένη χρονική στιγμή.Ο μοναδικός σκοπός του πράκτορα είναι να μεγιστοποιήσει τη συνολική ανταμοιβή του μέσα σε ένα επεισόδιο, δηλαδή οτιδήποτε συμβαίνει μεταξύ της αρχικής και της τελικής κατάστασης μέσα στο περιβάλλον.Ενισχύουμε τον πράκτορα ώστε να μάθει να εκτελεί τις καλύτερες ενέργειες μέσω της εμπειρίας. Αυτή είναι η πολιτική. Η πολιτική μπορεί να είναι μια απλή συνάρτηση ή ένας πίνακας αναζήτησης, ενώ σε άλλες περιπτώσεις μπορεί να περιλαμβάνει εκτεταμένους υπολογισμούς, όπως μια διαδικασία αναζήτησης. Η πολιτική είναι ο πυρήνας ενός πράκτορα ενισχυτικής μάθησης, με την έννοια ότι από μόνη της αρκεί για να καθορίσει τη συμπεριφορά.

• **Η συνάρτηση επιβράβευσης**, η οποία καθορίζει τον στόχο του συστήματος αναθέτοντας έναν αριθμό σε κάθε κατάσταση ή σε κάθε ζεύγος κατάστασης-ενέργειας (ζεύγος τιμής-ενέργειας) του πράκτορα, ανάλογα με τον αλγόριθμο που χρησιμοποιείται. Καθώς ο σκοπός του πράκτορα είναι να μεγιστοποιήσει μακροπρόθεσμα το άθροισμα των επιβραβεύσεων που λαμβάνει, η συνάρτηση επιβράβευσης καθορίζει ποιες καταστάσεις ή ποιοι συνδυασμοί καταστάσεων-ενεργειών είναι επιθυμητές.

• **Η συνάρτηση αξίας**, παρόμοια με τη συνάρτηση επιβράβευσης, αντιστοιχίζει κάθε κατάσταση (ή κατάσταση-ενέργεια) σε μία αριθμητική τιμή.

Η διαφορά είναι πως ενώ η συνάρτηση επιβράβευσης αφορά την άμεση επιβράβευση που λαμβάνει ο πράκτορας όταν φτάσει την κάθε κατάσταση, η συνάρτηση αξίας λαμβάνει υπόψιν την και τις μελλοντικές καταστάσεις στις οποίες είναι πιθανό να μεταβεί ο πράκτορας από την παρούσα κατάσταση και τις επιβραβεύσεις που θα λάβει τότε. Με αυτόν τον τρόπο η συνάρτηση αξίας εκφράζει τη μακροπρόθεσμη επιθυμητότητα κάθε κατάστασης.

- **Το μοντέλο του περιβάλλοντος**, το οποίο χρησιμοποιεί ο πράκτορας για να προβλέψει πώς οι ενέργειές του θα επηρεάσουν το περιβάλλον.Αυτό το στοιχείο δεν είναι κοινό σε όλους τους αλγορίθμους ενισχυτικής. Αντίθετα, πρόκειται για μία σχετικά νέα προσθήκη στο πεδίο, καθώς τα πρώτα συστήματα ενισχυτικής μάθησης βασίζονταν αποκλειστικά σε μεθόδους δοκιμής και σφάλματος.

Υπάρχουν δύο τύποι αλγορίθμων ενισχυτικής μάθησης. Είναι οι **βασισμένοι σε μοντέλα (model-based)** και οι **χωρίς μοντέλα (model-free)**.

Ένας model-free αλγόριθμος είναι ένας αλγόριθμος που εκτιμά τη βέλτιστη πολιτική χωρίς να χρησιμοποιεί συναρτήσεις μετάβασης και ανταμοιβής του περιβάλλοντος. Ενώ, ένας model-based αλγόριθμος είναι ένας αλγόριθμος που χρησιμοποιεί τη συνάρτηση μετάβασης και τη συνάρτηση ανταμοιβής προκειμένου να εκτιμήσει τη βέλτιστη πολιτική.

4.2 MDP (Markov Decision Process)

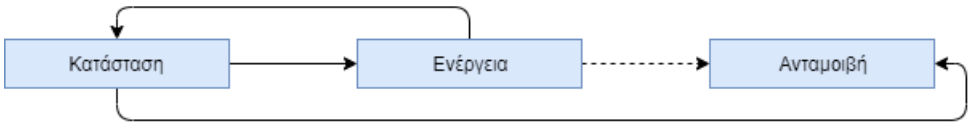

Figure 4.2: Q-learning
"Τροποποιήθηκε από την πηγή: [29]"

Η ενισχυτική μάθηση είναι ένας κλάδος της μηχανικής μάθησης αλγορίθμων όπου ένας πράκτορας αλληλεπιδρά με ένα περιβάλλον και μαθαίνει τη λύση ενός προβλήματος μέσω δοκιμής και σφάλματος.Σε μια μεγάλη πλειοψηφία των περιπτώσεων θεωρείται ότι το περιβάλλον είναι μαρκοβιανό το οποίο σημαίνει ότι ισχύει η ιδιότητα του Markov.Το πρόβλημα προσομοιώνεται σε ένα διακριτό χρονοδιάγραμμα σε ένα περιβάλλον μηχανικής

μάθησης.Αυτό σημαίνει ότι η μετάβαση στην επόμενη κατάσταση (state) μετά από την πραγματοποίηση μιας ενέργειας (action) εξαρτάται μόνο από την ενέργεια (action) αυτή και την παρούσα κατάσταση (state) στην οποία βρίσκεται ο πράκτορας (agent).Δεν υπάρχει δηλαδή εξάρτηση από τις προηγούμενες καταστάσεις (states) στις οποίες βρέθηκε ο πράκτορας.Αυτή η ιδιότητα, ακόμα και όταν ισχύει κατά προσέγγιση, απλοποιεί πολύ το μοντέλο του περιβάλλοντος και διευκολύνει τους υπολογισμούς με σχετικά απλούς αλγορίθμους [30].

Σε κάθε χρονικό βήμα t, ο πράκτορας βρίσκεται σε μια κατάσταση st S και αλληλεπιδρά με το περιβάλλον αναλαμβάνοντας μια ενέργεια a_t A(s).Στη συνέχεια, ο πράκτορας λαμβάνει μια νέα παρατήρηση του περιβάλλοντος και μετακινείται σε μια νέα κατάσταση st + 1. Επιπλέον, ο πράκτορας λαμβάνει ένα σήμα ανταμοιβής που δείχνει πόσο καλή είναι ή ήταν η ενέργεια που εκτέλεσε για την κατάσταση αυτή.

Η ιδιότητα αυτή μπορεί να γραφτεί ως εξής:

$P\{S_{t+1} = s', r_{t+1} = r'|S_t, a_t, r_t, S_{t-1}, a_{t-1}, r_{t-1}, ...\} = P\{S_{t+1} = s', r_{t+1} = r'|S_t, a_t, r_t\}$

Όπου το πρώτο μέλος της εξίσωσης εκφράζει την πιθανότητα να μεταβεί ο πράκτορας (agent) στην κατάσταση (state) s' και να λάβει ανταμοιβή (reward) r' δεδομένου ότι τη χρονική στιγμή t βρίσκεται στην κατάσταση S_t, πραγματοποιεί την ενέργεια (action) a_t και λαμβάνει ανταμοιβή r_t, τη χρονική στιγμή $t-1$ βρισκόταν στην κατάσταση S_{t-1} πραγματοποίησε την ενέργεια a_{t-1} και έλαβε επιβράβευση r_{t-1} κ.ο.κ.

Ενώ το δεύτερο μέλος δηλώνει ότι αυτή η πιθανότητα είναι ίση με το να μεταβεί ο πράκτορας στην κατάσταση s' και να λάβει επιβράβευση r' δεδομένου απλά ότι τη χρονική στιγμή t βρίσκεται στην κατάσταση S_t, πραγματοποιεί την ενέργεια a_t και λαμβάνει reward r_t.

4.3 Μελέτη αλγορίθμου Q-learning

Μία από τις βασικότερες τεχνικές της ενισχυτικής μάθησης, αποτελεί η τεχνική του Q-learning η οποία μπορεί να χρησιμοποιηθεί για την εύρεση της βέλτιστης πολιτικής π*, όταν δεν είναι εφικτή η εφαρμογή αλγορίθμων δυναμικού προγραμματισμού [31] [32].

$$\mathrm{Q(s,a)} = (1 - \mathrm{a})\mathrm{Q(s,a)} + \mathrm{a}\left(\mathrm{r} + \gamma \max\left(\mathrm{Q}\left(\mathrm{s}', \mathrm{a}'\right)\right)\right)$$

Ο Q-learning είναι ένας αλγόριθμος ενισχυτικής μάθησης εκτός πολιτικής (model-free) που επιδιώκει να βρει την καλύτερη ενέργεια για να πραγματοποιήσει, λαμβάνοντας υπόψη την τρέχουσα κατάσταση.Ο αλγόριθμος αυτός θεωρείται εκτός πολιτικής, γιατί η λειτουργία q -learning αντλεί δεδομένα και πληροφορίες από δράσεις που βρίσκονται εκτός της τρέχουσας πολιτικής, όπως η λήψη τυχαίων ενεργειών με αποτέλεσμα να μην είναι απαραίτητη μια καθορισμένη στρατηγική.Πιο συγκεκριμένα, ο αλγόριθμος q – learning προσπαθεί να μάθει μια πολιτική που να μεγιστοποιεί τη συνολική ανταμοιβή [33].

Για να εφαρμοστεί ο αλγόριθμος Q-learning συνήθως χρησιμοποιείται το πεπερασμένο μοντέλο της Μαρκοβιανής διεργασίας απόφασης (Finite Markov Decision Process - FMDP).Μέσω αυτής της διεργασίας, ο αλγόριθμος επιχειρεί να βρει τη βέλτιστη στρατηγική ώστε να επιτευχθεί η μεγιστοποίηση της προσδοκώμενης αξίας της συνολικής ανταμοιβής, ξεκινώντας από μια αρχική κατάσταση και πηγαίνοντας σε όλα τα διαδοχικά στάδια [27].

Ο αλγόριθμος Q-learning είναι ένας αλγόριθμος ενισχυτικής μάθησης που περιλαμβάνει έναν πράκτορα, ένα σύνολο καταστάσεων S, και ένα σύνολο ενεργειών A που μπορούν να εμφανιστούν σε κάθε κατάσταση.Ο πράκτορας εκτελώντας μια ενέργεια α που ανήκει στο σύνολο των ενεργειών A, μεταβαίνει από μια κατάσταση σε μια άλλη κατάσταση.Εκτελώντας μια ενέργεια σε μια συγκεκριμένη κατάσταση, ο πράκτορας ανταμείβεται με μια ανταμοιβή.

Μετά από κάποια βήματα Δt στο μέλλον, ο πράκτορας θα πάρει την απόφαση να πραγματοποιήσει ένα επόμενο του βήμα.Το βάρος (αξία) του συγκεκριμένου βήματος υπολογίζεται ως $\gamma^{\Delta}t$, όπου το γ (συντελεστής προεξόφλησης) είναι ένας αριθμός ανάμεσα στο μηδέν και στο ένα ($0 \leq \gamma \leq 1$) και έχει την τάση να αξιολογεί τις ανταμοιβές που λαμβάνονται νωρίτερα, υψηλότερα από εκείνες που λαμβάνονται αργότερα.Η τάση αυτή αντικατοπτρίζει την αξία μιας καλής εκκίνησης με άλλα λόγια.Το γ επίσης, μπορεί να ερμηνευτεί και ως η πιθανότητα της επιτυχίας ή της επιβίωσης σε κάθε βήμα Δt.

Για αυτό το λόγο, ο αλγόριθμος διαθέτει μια συγκεκριμένη συνάρτηση που υπολογίζει την ποιότητα του συνδυασμού μιας κατάστασης και μιας ενέργειας:

$Q : S \times A \rightarrow R$

Πριν ξεκινήσει η διαδικασία της μάθησης, ο πίνακας **Q** έχει μια αρχική τιμή που ορίζεται από τον προγραμματιστή.Σε κάθε χρονική στιγμή **t**, ο πράκτορας επιλέγει μια ενέργεια **at**, παρατηρεί μια ανταμοιβή **rt** και στο τέλος εισέρχεται σε μια καινούρια κατάσταση **St+1**, η οποία μπορεί να εξαρτάται είτε από την προηγούμενη κατάσταση **St** είτε από την ενέργεια που θα επιλέξει ο πράκτορας.Ως αποτέλεσμα όλων των προηγουμένων ο πίνακας **Q** ενημερώνεται με βάση τις παρατηρήσεις και τις ανταμοιβές που λαμβάνει ο πράκτορας.

Η κεντρική ιδέα αυτού του αλγορίθμου είναι η εφαρμογή της εξίσωσης Bellman, η οποία ανανεώνει μια τιμή με απλή επανάληψη.Η εξίσωση αυτή χρησιμοποιεί τον σταθμισμένο μέσο όρο των καινούριων πληροφοριών και της παλιάς αξίας.

$$Q^{new}\left(s_t, a_t\right) \leftarrow Q\left(\ \mathrm{s}_t, a_t\right) + \mathrm{a} \times \left(\mathrm{r}_t + \gamma \times \max Q\left(\ \mathrm{s}_{t+1}, \mathrm{a}\right) - \mathrm{Q}\left(s_t, a_t\right)\right)$$

Παραπάνω δίνεται η εξίσωση της Q-learning, όπου το r_t συμβολίζει την ανταμοιβή που λαμβάνει ο πράκτορας από την μεταβίβαση του από την κατάσταση s_t στην κατάσταση s_{t+1}, ενώ το a αντικατοπτρίζει το ρυθμό μάθησης που καθορίζει σε ποιο βαθμό οι νέες πληροφορίες υπερισχύουν των παλιών και κυμαίνεται $0 < a \leq 1$.

- Q $(s_t,\ a_t)$ = παλιά τιμή (old value)
- a = ρυθμός μάθησης
- r_t = ανταμοιβή (reward)
- γ = συντελεστή προεξόφλησης (discount factor)
- $\max Q\left(s_{t+1}, a\right)$ = εκτίμηση της βέλτιστης μελλοντικής αξίας
- $\left(r_t + \gamma \times \max Q\left(\ \mathrm{s}_{t+1}, \mathrm{a}\right) - \mathrm{Q}\left(s_t, a_t\right)\right)$ = χρονική διαφορά
- $r_t + \gamma \times \max Q\left(\ \mathrm{s}_{t+1}, \mathrm{a}\right)$ = νέα τιμή (στόχος χρονικής διαφοράς)

Συμπερασματικά, νέο Q (Qnew) είναι το άθροισμα τριών παραγόντων:

- Την τρέχουσα τιμή σταθμισμένη από το ρυθμό μάθησης

$(1-\alpha)Q(s_t, a_t)$. Συμπερασματικά οι τιμές του ρυθμού μάθησης που είναι κοντά στο 1, προκαλούν ταχύτερες αλλαγές στην τιμή του Q.

•Το γινόμενο $a \times r_t$, η ανταμοιβή δηλαδή r_t που λαμβάνεται όταν εκτελείται μια ενέργεια a_t σε μία κατάσταση S_t.

• Τέλος, τη μέγιστη ανταμοιβή που μπορεί να ληφθεί από κάποια επόμενη κατάσταση $max(Q(St+1, a))$.

Ένας πλήρης κύκλος του αλγορίθμου τερματίζει όταν η κατάσταση S_{t+1} είναι η τελική ή η τερματική κατάσταση.Μέσω του αλγορίθμου Q-learning, υπάρχει η δυνατότητα της μάθησης να είναι υλοποιήσιμη σε ατέρμονες εργασίες αν ο παράγοντας προεξόφλησης είναι μικρότερος της μονάδας.Με αποτέλεσμα οι αξίες των ενεργειών να παραμένουν πεπερασμένες ακόμα και αν το πρόβλημα περιέχει άπειρες επαναλήψεις.

Για όλες τις τελικές καταστάσεις S_f, $Q(S_f, a)$ δεν πραγματοποιείται ποτέ ενημέρωση στο Q, αλλά ορίζεται στη τιμή της ανταμοιβής r που παρατηρείται για την τελική κατάσταση S_f. Τέλος, σε πολλές περιπτώσεις το Q μπορεί να γίνει και ίσο με το μηδέν.

Βήμα Βήμα Υπολογισμός των τιμών του Q-learning αλγορίθμου

Στην εξίσωση Bellman συνήθως προσθέτουμε έναν συντελεστή βαρύτητας λ (learning rate), ο οποίος σταθμίζει την άμεση και την μελλοντική ανταμοιβή, αντίστοιχα, όπως φαίνεται παρακάτω:

$$Q(s_t, a_t) = (1-\lambda)Q(s_t, a_t) + \lambda * \left(R_t + \gamma \max_a Q(s_{t+1}, a)\right), \tag{4.1}$$

$$Q(s_t, a_t) \leftarrow (1-\alpha) \cdot \underbrace{Q(s_t, a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \overbrace{\left(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}}\right)}^{\text{learned value}}$$

Figure 4.3: Εξίσωση Bellman.
"Τροποποιήθηκε από την πηγή: https://en.wikipedia.org/wiki/Q-learning"

Η παραπάνω εξίσωση μας επιτρέπει να ενημερώνουμε σταδιακά τον πίνακα Q (Quality table ή Q-table) μετά από κάθε βήμα εκτέλεσης του αλγορίθμου Q-learning. Ο δισδιάστατος πίνακας τιμών ποιότητας Q , κρατά της αξία (Q-value) κάθε ενέργειας a δεδομένης μιας κατάστασης s. Σύμφωνα με την παραπάνω εξίσωση, ο ευφυής πράκτορας υπολογίζει την ανταμοιβή της τρέχουσας χρονικής στιγμής t με βάσει την προσδωκόμενη βέλτιστη μελλοντική ανταμοιβή της χρονικής στιγμής $t + 1$. Η βασική παραδοχή που κάνουμε είναι ότι ο ευφυής πράκτορας επιλέγει σε κάθε κατάσταση s πάντα την καλύτερη τρέχουσα γνωστή ενέργεια a. Σε ένα σύστημα συστάσεων, ο ευφυής πράκτορας θα αναζητήσει όλα τα δυνατά υποψήφια προς σύσταση στοιχεία για το προφίλ ενός συγκεκριμένου χρήστη και θα επιλέξει το ζεύγος χρήστη-στοιχείο (κατάσταση-ενέργεια) με την υψηλότερη αντίστοιχη τιμή Q.

Algorithm 1

Require:
States $S = \{1, ..., n_s\}$
Actions $A = \{1, ..., n_a\}$,
Reward function $R : S \times A \to R$
Transition probability function $T : S \times A \to S$
Learning rate $\lambda \in [0, 1]$, typically $\lambda = 0.1$
Discounting factor $\gamma \in [0, 1]$
Ensure: A Q-table which holds the value of each action a at state s.

procedure Q-LEARNING($S, A, R, T, \lambda, \gamma$)
  Initialize Q-table : $S \times A \to R$
  while Q-table is not converged **do**
    Start in state $s \in S$
    while s is not terminal **do**
      $\pi(s) \leftarrow argmax_a Q(s, a))$ ▷ compute policy π
      $a \leftarrow \pi(s)$ ▷ $\pi : S \to A$
      $r \leftarrow R(s, a)$ ▷ Receive the reward
      $s' \leftarrow T(s, a)$ ▷ Receive the new state
      $Q(s', a) \leftarrow (1 - \lambda) \times Q(s, a) + \lambda \times (r + \gamma \times max_{a'} Q(s', a'))$
      $s \leftarrow s'$
    end while
  end while
return Q
end procedure

Για να διευκολύνουμε την περιγραφή του αλγορίθμου Q-learning, θα χρησιμοποιήσουμε το παράδειγμα του παρακάτω πίνακα βαθμολογίας χρηστών

και στοιχείων, όπου U_{0-5} είναι χρήστες, ενώ I_{0-5} είναι τα στοιχεία που αυτοί έχουν βαθμολογήσει. Η κλίμακα βαθμολογίας είναι 0 έως 5 βαθμοί.

	I_0	I_1	I_2	I_3	I_4	I_5
U_0	5	1	1	4	0	0
U_1	1	5	4	0	0	4
U_2	2	1	5	0	0	4
U_3	1	2	1	5	0	0
U_4	5	1	2	0	1	0
U_5	1	5	4	0	4	0

Δισδιάστατος Πίνακας βαθμολογιών Χρηστών-Στοιχείων

Ο Q-table αρχικοποιείται με μηδενικά όπως φαίνεται στο αριστερό μέρος του Σχήματος 4.4.

	I_0	I_1	I_2	I_3	I_4	I_5
U_0	0	0	0	0	0	0
U_1	0	0	0	0	0	0
U_2	0	0	0	0	0	0
U_3	0	0	0	0	0	0
U_4	0	0	0	0	0	0
U_5	0	0	0	0	0	0

Αρχικοποίηση του πίνακα Q με μηδενικά

	I_0	I_1	I_2	I_3	I_4	I_5
U_0	0	0	0	0	0	0
U_1	0	0	0	0	0	0
U_2	0	0	0	0	0	0
U_3	0.25	0	0	0	0	0
U_4	0	0	0	0	0	0
U_5	0	0	0	0	0	0

Πίνακας Q μετά την 1η Επανάληψη

Figure 4.4: Ο Πίνακας με τις Q-values του παραδείγματος μας στην αρχικοποίηση του (αριστερά) και μετά την 1η επανάληψη (δεξιά).

Για την πρώτη επανάληψη:

Επιλέγουμε με τυχαίο τρόπο μια ενέργεια a (δηλαδή ένα προτεινόμενο στοιχείο i) για μια τυχαία κατάσταση s (δηλαδή ένα χρήστη u). Συγκεκριμένα, ο ευφυής πράκτορας προτείνει στον χρήστη με id = 3 (U_3) το στοιχείο με id = 0 (I_0), με ανταμοιβή r = 1. Η αξία-Q για την ενέργεια αυτή είναι Q(3,0) = 0.25. Οι ενημερώσεις των μεταβλητών του Αλγορίθμου 1 για την πρώτη επανάληψη είναι οι ακόλουθες:

$a \leftarrow 0$, $r \leftarrow 1$, $s' \leftarrow 3$,

$Q(s', a) \leftarrow (1 - \lambda) \times Q(s, a) + \lambda \times (r + \gamma \times max_{a'} Q(s', a')) \Rightarrow$
$\Rightarrow$Q(3,0)$\leftarrow 0.75 * 0 + 0.25 * (1 + 0.75 * 0) \Rightarrow$Q(3,0)$\leftarrow$ 0.25,

$s \leftarrow 3$

Συνεπώς, μετά το τέλος της 1ης επανάληψης, ο Q-table ενημερώνεται στο κελί $Q(3, 0)$ με την τιμή 0.25, όπως φαίνεται με κόκκινο χρώμα στο δεξιό μέρος του Σχήματος 4.4.

	I_0	I_1	I_2	I_3	I_4	I_5
U_0	0	0	0	0	0	0
U_1	0	0	0	0	0	0
U_2	0	0	0	0	0	0
U_3	0.25	0	0	0	0	0
U_4	1.25	0	0	0	0	0
U_5	0	0	0	0	0	0

Πίνακας Q μετά τη 2η Επανάληψη

	I_0	I_1	I_2	I_3	I_4	I_5
U_0	0	0	0	0	0	0
U_1	0	0	0	0	0	0
U_2	0	0	0	0	0	0
U_3	0.5	0	0	0	0	0
U_4	1.25	0	0	0	0	0
U_5	0	0	0	0	0	0

Πίνακας Q μετά τη 3η Επανάληψη

Για την δεύτερη επανάληψη:

Επιλέγουμε με τυχαίο τρόπο μια ενέργεια a (δηλαδή ένα προτεινόμενο στοιχείο i) για μια τυχαία κατάσταση s (δηλαδή ένα χρήστη u). Συγκεκριμένα, ο ευφυής πράκτορας προτείνει στον χρήστη με id = 4 (U_4) το στοιχείο με id = 0 (I_0), με ανταμοιβή r = 5. Η αξία-Q για την ενέργεια αυτή είναι Q(4,0) = 1.25. Οι ενημερώσεις των μεταβλητών του Αλγορίθμου 1 για την δεύτερη επανάληψη είναι οι ακόλουθες:

$a \leftarrow 0$, $r \leftarrow 5$, $s' \leftarrow 4$,

$Q(s', a) \leftarrow (1 - \lambda) \times Q(s, a) + \lambda \times (r + \gamma \times max_{a'} Q(s', a')) \Rightarrow$
$\Rightarrow$Q(4,0)$\leftarrow 0.75 * 0 + 0.25 * (5 + 0.75 * 0) \Rightarrow$Q(4,0)$\leftarrow$ 1.25,

$s \leftarrow 4$

Συνεπώς, μετά το τέλος της 2ης επανάληψης, ο Q-table ενημερώνεται στο κελί $Q(4, 0)$ με την τιμή 1.25, όπως φαίνεται με κόκκινο χρώμα στο αριστερό μέρος του Σχήματος 4.4.

Για την τρίτη επανάληψη:

Επιλέγουμε με τυχαίο τρόπο μια ενέργεια a (δηλαδή ένα προτεινόμενο στοιχείο i) για μια τυχαία κατάσταση s (δηλαδή ένα χρήστη u). Συγκεκριμένα,

ο ευφυής πράκτορας προτείνει στον χρήστη με id = 3 (U_3) το στοιχείο με id = 0 (I_0), με ανταμοιβή r = 1. Η αξία-Q για την ενέργεια αυτή είναι Q(3,0) = 0.5. Οι ενημερώσεις των μεταβλητών του Αλγορίθμου 1 για την τρίτη επανάληψη είναι οι ακόλουθες:

$a \leftarrow 0$, $r \leftarrow 1$, $s' \leftarrow 3$,

$Q(s', a) \leftarrow (1 - \lambda) \times Q(s, a) + \lambda \times (r + \gamma \times max_{a'} Q(s', a')) \Rightarrow$
$\Rightarrow$Q(3,0)$\leftarrow 0.75 * 0.25 + 0.25 * (1 + 0.75 * 0.25) \Rightarrow$Q(3,0)$\leftarrow$ 0.5,

$s \leftarrow 3$

Συνεπώς, μετά το τέλος της 3ης επανάληψης, ο Q-table ενημερώνεται στο κελί $Q(3, 0)$ με την τιμή 0.5, όπως φαίνεται με κόκκινο χρώμα στο δεξιό μέρος του Σχήματος 4.4.

	I_0	I_1	I_2	I_3	I_4	I_5
U_0	0	0	0	0	0	0
U_1	0	0	0	0	0	0
U_2	0	0	0	0	0	0
U_3	0.5	0	0	0	0	0
U_4	2.42	0	0	0	0	0
U_5	0	0	0	0	0	0

	I_0	I_1	I_2	I_3	I_4	I_5
U_0	0	0	0	0	0	0
U_1	0.25	0	0	0	0	0
U_2	0	0	0	0	0	0
U_3	0.5	0	0	0	0	0
U_4	2.42	0	0	0	0	0
U_5	0	0	0	0	0	0

Πίνακας Q μετά τη 4η Επανάληψη Πίνακας Q μετά τη 5η Επανάληψη

Για την τέταρτη επανάληψη:

Επιλέγουμε με τυχαίο τρόπο μια ενέργεια a (δηλαδή ένα προτεινόμενο στοιχείο i) για μια τυχαία κατάσταση s (δηλαδή ένα χρήστη u). Συγκεκριμένα, ο ευφυής πράκτορας προτείνει στον χρήστη με id = 4 (U_4) το στοιχείο με id = 0 (I_0), με ανταμοιβή r = 5. Η αξία-Q για την ενέργεια αυτή είναι Q(4,0) = 2.42. Οι ενημερώσεις των μεταβλητών του Αλγορίθμου 1 για την τέταρτη επανάληψη είναι οι ακόλουθες:

$a \leftarrow 0$, $r \leftarrow 5$, $s' \leftarrow 4$,

$Q(s', a) \leftarrow (1 - \lambda) \times Q(s, a) + \lambda \times (r + \gamma \times max_{a'} Q(s', a')) \Rightarrow$
$\Rightarrow$Q(4,0)$\leftarrow 0.75 * 1.25 + 0.25 * (1 + 0.75 * 0.25) \Rightarrow$Q(4,0)$\leftarrow$ 2.42,

$s \leftarrow 4$

Συνεπώς, μετά το τέλος της 4ης επανάληψης, ο Q-table ενημερώνεται στο

κελί $Q(4,0)$ με την τιμή 2.42, όπως φαίνεται με κόκκινο χρώμα στο αριστερό μέρος του Σχήματος 4.4.

Για την πέμπτη επανάληψη:

Επιλέγουμε με τυχαίο τρόπο μια ενέργεια a (δηλαδή ένα προτεινόμενο στοιχείο i) για μια τυχαία κατάσταση s (δηλαδή ένα χρήστη u). Συγκεκριμένα, ο ευφυής πράκτορας προτείνει στον χρήστη με id = 1 (U_1) το στοιχείο με id = 0 (I_0), με ανταμοιβή r = 1. Η αξία-Q για την ενέργεια αυτή είναι Q(1,0) = 0.25. Οι ενημερώσεις των μεταβλητών του Αλγορίθμου 1 για την πέμπτη επανάληψη είναι οι ακόλουθες:

$a \leftarrow 0$, $r \leftarrow 1$, $s' \leftarrow 1$,

$Q(s',a) \leftarrow (1-\lambda) \times Q(s,a) + \lambda \times (r + \gamma \times max_{a'}\, Q(s',a')) \Rightarrow$
$\Rightarrow$Q(1,0)$\leftarrow 0.75 * 0 + 0.25 * (1 + 0.75 * 0) \Rightarrow$Q(1,0)$\leftarrow$ 0.25,

$s \leftarrow 1$

Συνεπώς, μετά το τέλος της 5ης επανάληψης, ο Q-table ενημερώνεται στο κελί $Q(1,0)$ με την τιμή 0.25, όπως φαίνεται με κόκκινο χρώμα στο δεξιό μέρος του Σχήματος 4.4.

	I_0	I_1	I_2	I_3	I_4	I_5
U_0	0	0	0	0	0	0
U_1	0.25	0	0	0	0	0
U_2	0.5	0	0	0	0	0
U_3	0.5	0	0	0	0	0
U_4	2.42	0	0	0	0	0
U_5	0	0	0	0	0	0

Πίνακας Q μετά τη 6η Επανάληψη

	I_0	I_1	I_2	I_3	I_4	I_5
U_0	0	0	0	0	0	0
U_1	0.25	0	0	0	0	0
U_2	0.5	0	0	0	0	0
U_3	0.5	0	0	1.34	0	0
U_4	2.42	0	0	0	0	0
U_5	0	0	0	0	0	0

Πίνακας Q μετά τη 7η Επανάληψη

Για την έκτη επανάληψη:

Επιλέγουμε με τυχαίο τρόπο μια ενέργεια a (δηλαδή ένα προτεινόμενο στοιχείο i) για μια τυχαία κατάσταση s (δηλαδή ένα χρήστη u). Συγκεκριμένα, ο ευφυής πράκτορας προτείνει στον χρήστη με id = 2 (U_2) το στοιχείο με id = 0 (I_0), με ανταμοιβή r = 2. Η αξία-Q για την ενέργεια αυτή είναι Q(2,0) = 0.5. Οι ενημερώσεις των μεταβλητών του Αλγορίθμου 1 για την έκτη επανάληψη είναι οι ακόλουθες:

$a \leftarrow 0$, $r \leftarrow 2$, $s' \leftarrow 2$,

$Q(s', a) \leftarrow (1 - \lambda) \times Q(s, a) + \lambda \times (r + \gamma \times max_{a'} Q(s', a')) \Rightarrow$
$\Rightarrow$Q(2,0)$\leftarrow 0.75 * 0 + 0.25 * (2 + 0.75 * 0) \Rightarrow$Q(2,0)$\leftarrow$ 0.5,

$s \leftarrow 2$

Συνεπώς, μετά το τέλος της 6ης επανάληψης, ο Q-table ενημερώνεται στο κελί $Q(2, 0)$ με την τιμή 0.5, όπως φαίνεται με κόκκινο χρώμα στο αριστερό μέρος του Σχήματος 4.4.

Για την έβδομη επανάληψη:

Επιλέγουμε με τυχαίο τρόπο μια ενέργεια a (δηλαδή ένα προτεινόμενο στοιχείο i) για μια τυχαία κατάσταση s (δηλαδή ένα χρήστη u). Συγκεκριμένα, ο ευφυής πράκτορας προτείνει στον χρήστη με id = 3 (U_3) το στοιχείο με id = 0 (I_3), με ανταμοιβή r = 5. Η αξία-Q για την ενέργεια αυτή είναι Q(3,3) = 1.34. Οι ενημερώσεις των μεταβλητών του Αλγορίθμου 1 για την έβδομη επανάληψη είναι οι ακόλουθες:

$a \leftarrow 3$, $r \leftarrow 5$, $s' \leftarrow 3$,

$Q(s', a) \leftarrow (1 - \lambda) \times Q(s, a) + \lambda \times (r + \gamma \times max_{a'} Q(s', a')) \Rightarrow$
$\Rightarrow$Q(3,3)$\leftarrow 0.75 * 0 + 0.25 * (5 + 0.75 * 0) \Rightarrow$Q(3,3)$\leftarrow$ 1.34,

$s \leftarrow 3$

Συνεπώς, μετά το τέλος της 7ης επανάληψης, ο Q-table ενημερώνεται στο κελί $Q(3, 3)$ με την τιμή 1.34, όπως φαίνεται με κόκκινο χρώμα στο δεξιό μέρος του Σχήματος 4.4.

4.4 Μελέτη αλγορίθμου Deep Q-learning Network (DQN)

Στη βαθιά εκμάθηση Q learning χρησιμοποιείται ένα νευρωνικό δίκτυο για τη προσέγγιση της συνάρτησης Q-value.Η κατάσταση (state) δίνεται ως είσοδος και η τιμή Q όλων των πιθανών ενεργειών (actions) παράγεται ως έξοδος.Η σύγκριση μεταξύ Q-learning και deep Q-learning απεικονίζεται παρακάτω [34]:

Ο αλγόριθμος Q-Learning ακολουθεί τα παρακάτω βήματα:

Figure 4.5: Q-learning
"Τροποποιήθηκε από την πηγή:[34]"

1. Αρχικοποίηση του πίνακα Q:

Ο πίνακας Q αντιστοιχίζει καταστάσεις και ενέργειες με τις αντίστοιχες τιμές Q, οι οποίες είναι εκτιμήσεις της βέλτιστης μελλοντικής τιμής. Ο πίνακας Q είναι μια απλή δομή δεδομένων που χρησιμοποιείται για την παρακολούθηση των καταστάσεων, των ενεργειών και των αναμενόμενων ανταμοιβών τους. Συγκεκριμένα, ο πίνακας Q αντιστοιχίζει ένα ζεύγος κατάστασης-ενέργειας σε μια τιμή Q, την οποία ο πράκτορας θα μάθει. Κατά την έναρξη του αλγορίθμου εκμάθησης Q, ο πίνακας Q αρχικοποιείται με όλα μηδενικά, υποδεικνύοντας ότι ο πράκτορας δεν γνωρίζει τίποτα για τον κόσμο. Καθώς ο πράκτορας δοκιμάζει διάφορες ενέργειες σε διαφορετικές καταστάσεις μέσω δοκιμής και σφάλματος, ο πράκτορας μαθαίνει την αναμενόμενη ανταμοιβή κάθε ζεύγους κατάστασης-ενέργειας και ενημερώνει τον πίνακα Q με τη νέα τιμή Q. Η διαδικασία της χρήσης της δοκιμής και του σφάλματος για τη μάθηση του κόσμου ονομάζεται εξερεύνηση.

Ένας από τους στόχους του αλγορίθμου Q-Learning είναι η εκμάθηση της τιμής Q για ένα νέο περιβάλλον.Η Q-Value είναι η μέγιστη αναμενόμενη ανταμοιβή που μπορεί να επιτύχει ένας πράκτορας λαμβάνοντας μια δεδομένη ενέργεια A από την κατάσταση S. Μόλις ένας πράκτορας μάθει την Q-Value για κάθε ζεύγους κατάστασης-ενέργειας, ο πράκτορας μεγιστοποιεί την αναμενόμενη ανταμοιβή του στην κατάσταση S επιλέγοντας την ενέργεια A με την υψηλότερη αναμενόμενη ανταμοιβή. Η ρητή επιλογή της καλύτερης γνωστής ενέργειας είναι γνωστή ως εκμετάλλευση.

2. Επιλογή μιας ενέργειας επιλέγοντας την στρατιγική Epsilon Greedy Exploration

- Σε κάθε χρονικό βήμα, όταν έρθει η ώρα να επιλέξουμε μια ενέργεια,

ρίχνουμε ένα ζάρι.

• Αν το ζάρι έχει πιθανότητα μικρότερη από το έψιλον, επιλέγουμε μια τυχαία ενέργεια.

• Διαφορετικά, επιλέγουμε την καλύτερη γνωστή ενέργεια στην τρέχουσα κατάσταση του πράκτορα.

Στην αρχή του αλγορίθμου, κάθε βήμα που κάνει ο πράκτορας θα είναι τυχαίο, κάτι που είναι χρήσιμο για να βοηθήσει τον πράκτορα να μάθει το περιβάλλον στο οποίο βρίσκεται.Καθώς ο πράκτορας κάνει όλο και περισσότερα βήματα, η τιμή του epsilon μειώνεται και ο πράκτορας αρχίζει να δοκιμάζει όλο και περισσότερες από τις υπάρχουσες γνωστές καλές ενέργειες.Το epsilon είναι αρχικοποιημένο στο 1, πράγμα που σημαίνει ότι κάθε βήμα είναι τυχαίο στην αρχή.Κοντά στο τέλος της διαδικασίας εκπαίδευσης, ο πράκτορας θα εξερευνά πολύ λιγότερο και θα εκμεταλλεύεται πολύ περισσότερο.

3. Ενημέρωση του πίνακα Q χρησιμοποιώντας την εξίσωση Bellman

Η εξίσωση Bellman μας λέει πώς να ενημερώνουμε τον πίνακα Q μετά από κάθε βήμα που κάνουμε.Συνοπτικά, ο πράκτορας ενημερώνει την τρέχουσα αντιλαμβανόμενη αξία με την εκτιμώμενη βέλτιστη μελλοντική ανταμοιβή, υποθέτοντας ότι ο πράκτορας λαμβάνει την καλύτερη επί του παρόντος γνωστή ενέργεια. Σε μια εφαρμογή, ο πράκτορας αναζητά όλες τις πιθανές ενέργειες για μια δεδομένη κατάσταση και επιλέγει το ζεύγος κατάστασης-ενέργειας με την υψηλότερη αντίστοιχη τιμή Q.

Ο αλγόριθμος Deep Q-Learning:

• Αρχικοποίηση των κύριων νευρωνικών δικτύων και των νευρωνικών δικτύων στόχου

Στο Deep Q-Learning, αντικαθιστούμε τον κανονικό πίνακα Q με ένα νευρωνικό δίκτυο. Το νευρωνικό δίκτυο απεικονίζει τις καταστάσεις εισόδου σε ζεύγη (δράση, τιμή Q). Η διαδικασία μάθησης χρησιμοποιεί δύο νευρωνικά δίκτυα με την ίδια αρχιτεκτονική αλλά διαφορετικά βάρη. Κάθε N βήματα, τα βάρη από το κύριο δίκτυο αντιγράφονται στο δίκτυο-στόχο. Η χρήση και των δύο αυτών δικτύων οδηγεί σε μεγαλύτερη σταθερότητα στη διαδικασία μάθησης και βοηθά τον αλγόριθμο να μαθαίνει πιο αποτελεσματικά.

Πώς να αντιστοιχίσετε καταστάσεις σε ζεύγη (Action,Q-value):

Τα κύρια νευρωνικά δίκτυα και το νευρωνικό δίκτυο-στόχος αντιστοιχίζουν τις καταστάσεις εισόδου σε ένα ζεύγος (ενέργεια, τιμή q).Στην περίπτωση αυτή, κάθε κόμβος εξόδου (που αντιπροσωπεύει μια ενέργεια) περιέχει την τιμή q της δράσης ως αριθμό κινητής υποδιαστολής.Οι κόμβοι εξόδου δεν αναπαριστούν κατανομή πιθανότητας, επομένως δεν θα έχουν άθροισμα 1.

• Επιλογή μιας ενέργειας χρησιμοποιώντας τη στρατηγική εξερεύνησης Epsilon-Greedy

Στη στρατηγική Epsilon-Greedy Exploration, ο πράκτορας επιλέγει μια τυχαία ενέργεια με πιθανότητα epsilon και εκμεταλλεύεται την καλύτερη γνωστή ενέργεια με πιθανότητα 1 - epsilon.

Πώς βρίσκετε την πιο γνωστή ενέργεια από το δίκτυό σας;

Τόσο το κύριο μοντέλο όσο και το μοντέλο στόχου αντιστοιχίζουν τις καταστάσεις εισόδου σε ενέργειες εξόδου, οι οποίες αντιπροσωπεύουν την προβλεπόμενη τιμή Q του μοντέλου. Σε αυτή την περίπτωση, η ενέργεια με την υψηλότερη προβλεπόμενη τιμή Q είναι η πιο γνωστή ενέργεια σε αυτή την κατάσταση.

• Ενημέρωση των βαρών του δικτύου χρησιμοποιώντας την εξίσωση Bellman

Αφού επιλέξει μια ενέργεια, ήρθε η ώρα ο πράκτορας να εκτελέσει την ενέργεια και να ενημερώσει τα δίκτυα Main και Target σύμφωνα με την εξίσωση Bellman.Οι πράκτορες Deep Q-Learning χρησιμοποιούν την επανάληψη εμπειριών για να μάθουν για το περιβάλλον τους και να ενημερώσουν τα δίκτυα Main και Target.

Μόλις επιλεγεί μια ενέργεια, είναι ώρα για τον πράκτορα να εκτελέσει την ενέργεια και να ενημερώσει τα δίκτυα Main και Target σύμφωνα με την εξίσωση Bellman. Οι πράκτορες Deep Q-Learning χρησιμοποιούν την επανάληψη εμπειριών για να μάθουν για το περιβάλλον τους και να ενημερώσουν τα δίκτυα Main και Target.

Συνοψίζοντας, το κύριο δίκτυο παίρνει δείγματα και εκπαιδεύεται σε μια

δέσμη προηγούμενων εμπειριών κάθε 4 βήματα.Στη συνέχεια, τα βάρη του κύριου δικτύου αντιγράφονται στα βάρη του δικτύου-στόχου κάθε 100 βήματα.

Επανάληψη εμπειρίας

Η αναπαραγωγή εμπειρίας είναι η πράξη της αποθήκευσης και αναπαραγωγής καταστάσεων του παιχνιδιού (κατάσταση, ενέργεια, ανταμοιβή, επόμενη κατάσταση) από τις οποίες ο αλγόριθμος RL μπορεί να μάθει.Η επανάληψη εμπειρίας μπορεί να χρησιμοποιηθεί σε αλγορίθμους εκτός πολιτικής για να μάθουν με τρόπο offline.Οι μέθοδοι εκτός πολιτικής (Off-Policy) είναι σε θέση να ενημερώνουν τις παραμέτρους του αλγορίθμου χρησιμοποιώντας αποθηκευμένες και αναπαραγόμενες πληροφορίες από προηγούμενες ενέργειες.Το Deep Q-Learning χρησιμοποιεί την αναπαραγωγή εμπειρίας (Experience Replay) για να μαθαίνει σε μικρές παρτίδες προκειμένου να αποφεύγεται η στρέβλωση της κατανομής του συνόλου δεδομένων των διαφορετικών καταστάσεων, ενεργειών, ανταμοιβών και επόμενων καταστάσεων που θα δει το νευρωνικό δίκτυο.Είναι σημαντικό ότι ο πράκτορας δεν χρειάζεται να εκπαιδεύεται μετά από κάθε βήμα.Στην υλοποίησή μας, χρησιμοποιούμε την επανάληψη εμπειρίας για να εκπαιδεύουμε σε μικρές παρτίδες μία φορά κάθε 4 βήματα αντί για μετά από κάθε βήμα.Διαπιστώσαμε ότι αυτό το τέχνασμα βοήθησε πραγματικά στην επιτάχυνση της υλοποίησης Deep Q-Learning.

Εξίσωση Bellman

Παραμένει απαραίτητο να πραγματοποιείται η ενημέρωση των βαρών του μοντέλου του πράκτορα σύμφωνα με την εξίσωση Bellman.

Από την αρχική εξίσωση Bellman στο Σχήμα 3.3, θέλουμε να αναπαραγάγουμε τη λειτουργία στόχου Temporal Difference χρησιμοποιώντας το νευρωνικό μας δίκτυο και όχι έναν πίνακα Q. Το δίκτυο-στόχος και όχι το κύριο δίκτυο χρησιμοποιείται για τον υπολογισμό του στόχου Temporal Difference. Υποθέτοντας ότι η λειτουργία στόχου temporal difference παράγει μια τιμή 9 στο παραπάνω παράδειγμα, μπορούμε να ενημερώσουμε τα βάρη του κύριου δικτύου αναθέτοντας 9 στην τιμή q-στόχου και προσαρμόζοντας τα βάρη του κύριου δικτύου μας στις νέες τιμές στόχου.

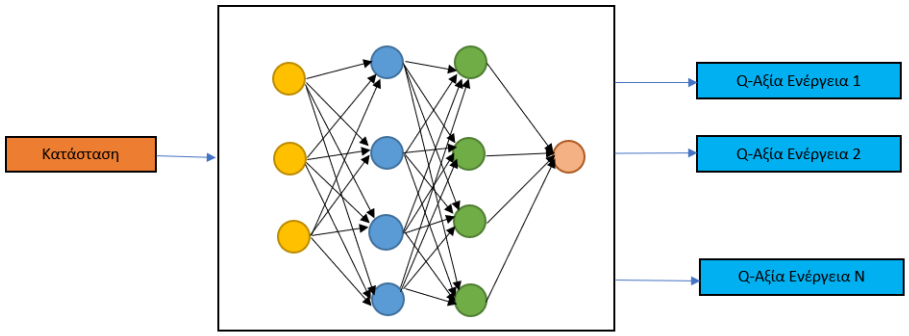

Figure 4.6: Deep Q-learning
"Τροποποιήθηκε από την πηγή: [34]"

4.5 Μελέτη αλγορίθμου Advantage Actor-Critic learning (A2C)

Στον τομέα της Ενισχυτικής Μάθησης, ο αλγόριθμος Advantage Actor Critic (A2C) συνδυάζει δύο διαφορετικούς τύπους αλγορίθμων Ενισχυτικής Μάθησης, τον πολιτικό (Policy Based) και τον βασισμένο σε αξίες (Value Based).

Οι πράκτορες που βασίζονται στην πολιτική μαθαίνουν απευθείας μια πολιτική (μια κατανομή πιθανότητας ενεργειών) που αντιστοιχίζει τις καταστάσεις εισόδου σε ενέργειες εξόδου.

• Οι μέθοδοι Q-learning και Deep Q-learning είναι βασισμένες στην αξία και λειτουργούν δημιουργώντας μια συνάρτηση αξίας που θα αντιστοιχίζει κάθε ζεύγος κατάστασης-ενέργειας σε μια τιμή. Με αυτές τις μεθόδους μπορούμε να βρούμε την καλύτερη ενέργεια για κάθε κατάσταση, δηλαδή την ενέργεια που έχει τη μεγαλύτερη τιμή. Αυτό είναι αποτελεσματικό όταν ο αριθμός των δυνατών ενεργειών είναι πεπερασμένος.

• Οι μέθοδοι βασισμένες στην πολιτική (REINFORCE με πολιτικές διαβαθμίσεις) βελτιστοποιούν απευθείας την πολιτική, χωρίς να χρησιμοποιούν μια συνάρτηση αξίας. Αυτό είναι χρήσιμο όταν ο χώρος των δράσεων είναι συνεχής ή στοχαστικός. Το κύριο πρόβλημα είναι η εύρεση μιας καλής συνάρτησης βαθμολογίας που να υπολογίζει πόσο καλή είναι μια πολιτική. Για αυτό χρησιμοποιούμε τις συνολικές ανταμοιβές ενός επεισοδίου.

Ωστόσο, και οι δύο αυτές μέθοδοι έχουν σημαντικά μειονεκτήματα. Για

αυτό το λόγο, θα αναλύσουμε μια νέα μέθοδο Ενισχυτικής Μάθησης, η οποία μπορεί να χαρακτηριστεί ως "υβριδική μέθοδος": Actor Critic. Θα χρησιμοποιήσουμε δύο νευρωνικά δίκτυα. [35]:

1.έναν κριτή που μετράει πόσο καλή είναι η ενέργεια που αναλαμβάνεται (με βάση την αξία) και

2.έναν πράκτορα που ελέγχει πώς συμπεριφέρεται ο πράκτοράς μας (βασισμένος στην πολιτική)

Τα νευρωνικά δίκτυα παραμετροποιούν τόσο τη λειτουργία Critic όσο και τη λειτουργία Actor. Η παραμετροποίηση της Critic λειτουργίας αφορά στην πρόβλεψη της τιμής Q, και συνεπώς αυτή η μέθοδος ονομάζεται Q-Actor Critic.

Η βέλτιστη βασική συνάρτηση είναι η συνάρτηση κατάστασης-αξίας. Χρησιμοποιώντας τη συνάρτηση V ως βασική συνάρτηση, αφαιρούμε την τιμή Q από την τιμή V. Αυτό μας δείχνει πόσο καλύτερα είναι να λάβουμε μια συγκεκριμένη ενέργεια σε σύγκριση με τη μέση γενική ενέργεια στη δεδομένη κατάσταση. Αυτή η τιμή ονομάζεται τιμή πλεονεκτήματος (advantage value):

$$A\left(s_t, a_t\right) = Q_w\left(s_t, a_t\right) - V_u\left(s_t\right)$$

Αυτό δεν σημαίνει ότι απαιτείται η κατασκευή δύο ξεχωριστών νευρωνικών δικτύων, ένα για την τιμή Q και ένα για την τιμή V. Μπορούμε αντίθετα να χρησιμοποιήσουμε τη σχέση μεταξύ Q και V που προκύπτει από την εξίσωση βελτιστότητας Bellman:

$$Q\left(s_t, a_t\right) = E\left[r_{t+1} + \gamma V\left(s_{t+1}\right)\right]$$

Έτσι, μπορούμε να ξαναγράψουμε το πλεονέκτημα ως εξής:

$$A\left(s_t, a_t\right) = r_{t+1} + \gamma V_u\left(s_{t+1}\right) - V_u\left(s_t\right)$$

Οπότε, χρειάζεται να χρησιμοποιήσουμε μόνο ένα νευρωνικό δίκτυο για τη συνάρτηση V (που παραμετροποιήθηκε από το v παραπάνω).Έτσι, μπορούμε να ξαναγράψουμε την εξίσωση ενημέρωσης ως εξής:

$$\nabla_\theta J(\theta) \sim \sum_{t=0}^{T-1} \nabla_\theta \log \pi_\theta \left(a_t \mid s_t\right) \left(r_{t+1} + \gamma V_u \left(s_{t+1}\right) - V_u \left(s_t\right)\right) = \sum_{t=0}^{T-1} \nabla_\theta \log \pi_\theta \left(a_t \mid s_t\right) A \left(s_t, a_t\right)$$

Αυτό είναι το Advantage Actor Critic.

Το Advantage Actor Critic έχει δύο κύριες παραλλαγές: το Asynchronous Advantage Actor Critic (A3C) και το Advantage Actor Critic (A2C).

Εφαρμογή του A2C:

Ως εκ τούτου, η νέα εξίσωση ενημέρωσης αντικαθιστά το προηγουμένως χρησιμοποιούμενο αθροιστικό σήμα ανταμοιβής με βάση τις κλίσεις πολιτικής βανίλιας με τη συνάρτηση Advantage.

$$\nabla_\theta J(\theta) \sim \sum_{t=0}^{T-1} \nabla_\theta \log \pi_\theta \left(a_t \mid s_t\right) A \left(s_t, a_t\right)$$

Σε κάθε βήμα μάθησης, ενημερώνουμε τόσο την παράμετρο Actor (χρησιμοποιώντας τις κλίσεις πολιτικής και την τιμή του πλεονεκτήματος) όσο και την παράμετρο Critic (χρησιμοποιώντας την εξίσωση ενημέρωσης Bellman για την ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος).

Επισκόπηση του αλγορίθμου Advantage Actor-Critic:

Ο αλγόριθμος actor critic αποτελείται από δύο δίκτυα (τον actor και τον critic) που συνεργάζονται για την επίλυση ενός συγκεκριμένου προβλήματος.Σε υψηλό επίπεδο, η Συνάρτηση Πλεονεκτήματος υπολογίζει το σφάλμα TD ή το σφάλμα πρόβλεψης του πράκτορα.Το δίκτυο actor επιλέγει μια ενέργεια σε κάθε χρονικό βήμα και το δίκτυο critic αξιολογεί την ποιότητα ή την τιμή Q μιας δεδομένης κατάστασης εισόδου.Καθώς το δίκτυο critic μαθαίνει ποιες καταστάσεις είναι καλύτερες ή χειρότερες, ο πράκτορας χρησιμοποιεί αυτές τις πληροφορίες για να μάθει στον εαυτό του να αναζητά καλές καταστάσεις και να αποφεύγει τις κακές.

Το πρόβλημα είναι ότι ο critic θα πρέπει να προσεγγίσει δύο συναρτήσεις: Qπ(s,a) και Vπ(s).

Η πρώτη πτυχή του A2C είναι ότι βασίζεται σε ενημέρωση n-βήματος, η οποία αποτελεί συμβιβασμό μεταξύ MC και TD:

Το MC περιμένει μέχρι το τέλος ενός επεισοδίου για να ανανεώσει την αξία μιας ενέργειας χρησιμοποιώντας την ανταμοιβή to-go (άθροισμα των αποκτηθέντων ανταμοιβών) R(s,a).

Το TD ενημερώνει αμέσως τη δράση χρησιμοποιώντας την άμεση ανταμοιβή r(s,a,s) και προσεγγίζει τα υπόλοιπα με την τιμή της επόμενης κατάστασης Vπ(s).

n-βήμα χρησιμοποιεί τις n επόμενες άμεσες ανταμοιβές και προσεγγίζει τις υπόλοιπες με την τιμή της κατάστασης που επισκέφθηκε n βήματα αργότερα.

Συνεπώς, ο TD μπορεί να θεωρηθεί ως αλγόριθμος ενός βήματος. Για αραιές ανταμοιβές (κυρίως μηδέν, +1 ή -1 στο τέλος ενός παιχνιδιού για παράδειγμα), αυτό επιτρέπει την ενημέρωση των n τελευταίων ενεργειών που οδηγούν σε νίκη/ήττα, αντί μόνο της τελευταίας στην TD, επιταχύνοντας τη μάθηση. Ωστόσο, δεν υπάρχει ανάγκη για πεπερασμένα επεισόδια όπως στο MC. Με άλλα λόγια, η εκτίμηση n-βήματος παρέχει ένα συμβιβασμό μεταξύ της μεροληψίας (εσφαλμένες ενημερώσεις βάσει εκτιμώμενων τιμών όπως στο TD) και της διακύμανσης (μεταβλητότητα των αποδόσεων που λαμβάνονται όπως στο MC). Μια εναλλακτική λύση στην ενημέρωση n-βήματος είναι η χρήση ιχνών επιλεξιμότητας.

Το A2C έχει μια αρχιτεκτονική κριτικής των φορέων.

Ο actor εξάγει την πολιτική πθ για μια κατάσταση s, δηλαδή ένα διάνυσμα πιθανοτήτων για κάθε ενέργεια.

Ο critic εξάγει την τιμή Vφ(s) μιας κατάστασης s.

Κεφάλαιο 5

Συστήματα Συστάσεων (Recommender Systems)

5.1 Εισαγωγή

Η αύξηση του όγκου δεδομένων στο διαδίκτυο στον 21ο αιώνα δημιούργησε πρόβλημα στη διαχείριση των πληροφοριών, καθώς το φαινόμενο της υπερπληροφόρησης καθιστά δύσκολη την άμεση πρόσβαση του χρήστη σε αυτές. Αυτό έχει ως αποτέλεσμα τη δυσκολία κατανόησης ενός προβλήματος και τη λήψη μιας σωστής απόφασης όταν ο όγκος των πληροφοριών είναι μεγάλος. Για αυτόν τον λόγο, απαιτούνται συστήματα φιλτραρίσματος δεδομένων που θα παρέχουν αποτελέσματα ανάλογα με τις ανάγκες του χρήστη. Έτσι, αναπτύχθηκαν συστήματα για το διαχωρισμό των πληροφοριών.

Τα Συστήματα Συστάσεων (Recommender Systems) είναι συστήματα φιλτραρίσματος πληροφοριών τα οποία αντιμετωπίζουν το πρόβλημα της υπερπληροφόρησης, φιλτράροντας τις σημαντικές πληροφορίες που προέρχονται από ένα μεγάλο όγκο παραγόμενων πληροφοριών από τις προτιμήσεις ή τη διαδικτυακή συμπεριφορά του χρήστη και εκφράζονται είτε άμεσα (explicitly) είτε έμμεσα (implicitly).Στόχος τους είναι να παρουσιάσουν μόνο τις πληροφορίες που ενδιαφέρουν πραγματικά τον κάθε χρήστη.

Μερικά παραδείγματα εφαρμογών που χρησιμοποιούν συστήματα συστάσεων είναι το Netflix για ταινίες, την Amazon για τις αγορές, τα

κοινωνικά δίκτυα Facebook και Twitter για να προτείνουν νέους φίλους και ομάδες και το GroupLens για άρθρα.

Σε όλες τις εφαρμογές, οι πελάτες έχουν πρόσβαση σε έναν μεγάλο όγκο πληροφοριών, και τα συστήματα συστάσεων τους βοηθούν να εντοπίσουν και να βρουν γρήγορα και εύκολα αυτό που πραγματικά θέλουν.

5.2 Διαδικασία Σύστασης

Κάθε σύστημα συστάσεων δέχεται μια είσοδο και στην συνέχεια ακολουθεί μια διαδικασία και παράγει κάποια αποτελέσματα.Τα στοιχεία που μπορεί ένα σύστημα συστάσεων να πάρει ως είσοδο προέρχονται είτε από τα δεδομένα του χρήστη, είτε από τα δεδομένα του αντικειμένου είτε από κάποιες αλληλεπιδράσεις μεταξύ του χρήστη με το αντικείμενο.

Κάθε χρήστης δημιουργεί ένα προφίλ στο οποίο έχει μέσα στοιχεία που θα τον βοηθήσουν να έχει σωστά αποτελέσματα από τα συστήματα συστάσεων. Ομοίως ένα αντικείμενο έχει κάποια χαρακτηριστικά τα οποία συμπληρώνουν ένα προφίλ σχετικά με το αντικείμενο, και χρησιμοποιούνται για να μπορούν να δίνουν την αντιστοίχηση από τα συστήματα συστάσεων.Τα δεδομένα στο προφίλ του χρήστη μπορούν να εισαχθούν είτε άμεσα είτε έμμεσα.Άμεση είναι η διαδικασία κατά την οποία ο χρήστης βάζει ως δεδομένα στο σύστημα τα προσωπικά του στοιχεία και πληροφορίες για τα ενδιαφέροντά του.Έμμεση είναι η διαδικασία κατά την οποία το σύστημα αντλεί πληροφορίες για τον χρήστη από τις προσωπικές του σελίδες, είτε από τις πληροφορίες που παρέχει το προσωπικό του κινητό τηλέφωνο, είτε από το ιστορικό των αγορών που έχει κάνει[36].

5.3 Κατηγορίες Βασικών Μοντέλων Συστημάτων Συστάσεων

Τα συστήματα συστάσεων χωρίζονται στις παρακάτω κατηγορίες οι οποίες διαφέρουν στο τρόπο με τον οποίο γίνονται οι συστάσεις μεταξύ των χρηστών.

•Στα Συστήματα που είναι βασισμένα στη Συνεργασία(Collaborative Systems)

•Στα Συστήματα που είναι βασισμένα στο Περιεχομένου (Content Based)

•Στα Συστήματα που είναι βασισμένα στη Γνώση(Knowledge- Based Systems)

Αυτές οι τρεις κατηγορίες αποτελούν τους θεμελιώδης πυλώνες της έρευνας στα συστήματα συστάσεων.

5.3.1 Συστήματα Βασισμένα στη Συνεργασία (Collaborative Systems)

Οι αλγόριθμοι συνεργατικού φιλτραρίσματος συλλέγουν πληροφορίες σχετικά με τις αγοραστικές συνήθειες ή προτιμήσεις των πελατών και παρέχουν συστάσεις σε άλλους χρήστες βασιζόμενοι στην ομοιότητα των συνολικών αγοραστικών προτύπων [37].Για να λειτουργήσουν αυτά τα συστήματα, απαιτείται η ύπαρξη αξιολογήσεων για αντικείμενα από άλλους χρήστες.

Συγκεκριμένα, για κάθε χρήστη εντοπίζεται ένα σύνολο πλησιέστερων χρηστών, γνωστών ως "γείτονες", με τους οποίους υπάρχει υψηλή συσχέτιση με βάση υπολογισμούς. Αυτό μας επιτρέπει να προβλέψουμε αποτελέσματα για άγνωστα στοιχεία, εκμεταλλευόμενοι τα αποτελέσματα που είναι γνωστά από τους πλησιέστερους "γείτονες". Οι αλγόριθμοι αυτοί συλλέγουν και αναλύουν δεδομένα σχετικά με τη δραστηριότητα του χρήστη και βασίζουν τις συστάσεις τους στις επιλογές άλλων χρηστών με παρόμοια συμπεριφορά.

Για παράδειγμα, οι πληροφορίες αποθηκεύονται σε έναν δισδιάστατο πίνακα, όπου η μία στήλη περιέχει τους χρήστες και η άλλη τις προτιμήσεις τους. Ο στόχος είναι να εντοπιστεί η "γειτονιά" του χρήστη, δηλαδή ένα υποσύνολο χρηστών που μοιράζονται κοινά χαρακτηριστικά με τον εξεταζόμενο χρήστη. Με βάση αυτήν τη γειτονιά, ο χρήστης μπορεί να λάβει προτάσεις για στοιχεία που δεν έχει αξιολογήσει ακόμα, αλλά ταιριάζουν στις προτιμήσεις του.[38] [39].

Είναι η δημοφιλέστερη και πλέον η πιο κυρίαρχη μέθοδος συστάσεων.Η συλλογή δεδομένων μπορεί να γίνει είτε άμεσα (explicitly) είτε έμμεσα (implicitly).

Η συλλογή δεδομένων μπορεί να γίνει **άμεσα** με διάφορους τρόπους.Ένας

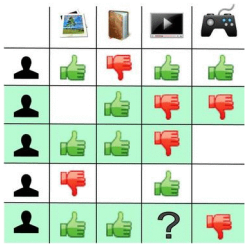

Figure 5.1: Αναπαράσταση μητρώου χρηστών
"Πηγή: Data Cadamia - rating collaborative filtering "

τρόπος είναι να ζητηθεί από τον χρήστη να βαθμολογήσει ένα αντικείμενο ή να ταξινομήσει μια λίστα αντικειμένων με σειρά προτίμησης.Επιπλέον, μπορεί να γίνει ρωτώντας τον χρήστη να επιλέξει ποιο αντικείμενο προτιμάει από ένα ζευγάρι αντικειμένων που του παρουσιάζονται.

Ενώ η **έμμεση** συλλογή δεδομένων μπορεί να γίνει είτε παρατηρώντας τα αντικείμενα που ο χρήστης επισκέπτεται ή αγοράζει όταν είναι συνδεδεμένος στον ιστότοπο, καθώς και πόσες φορές έχει επισκεφτεί ένα συγκεκριμένο αντικείμενο.Είτε παρατηρώντας τον κοικωνικό περίγυρο του χρήστη, προκειμένου να ανιχνευθούν παρόμοιες συμπεριφορές.

Κάποιες από τις χρήσεις του συνεργατικού φιλτραρίσματος είναι οι παρακάτω:

- Παροχή προσαρμοσμένων συστάσεων:

  Οι αλγόριθμοι συνεργατικού φιλτραρίσματος μπορούν να παρέχουν προσαρμοσμένες συστάσεις στους χρήστες βασιζόμενοι στις προτιμήσεις και τη συμπεριφορά άλλων παρόμοιων χρηστών.

- Προσωποποίηση εμπειρίας χρήστη:

  Οι αλγόριθμοι αυτοί μπορούν να βοηθήσουν στην προσαρμογή και εξατομίκευση της εμπειρίας των χρηστών, παρέχοντας τους συστάσεις που ανταποκρίνονται στις προτιμήσεις και τα ενδιαφέροντά τους. [40].

- Φιλτράρισμα πληροφορίας:

  Οι αλγόριθμοι αυτοί μπορούν να βοηθήσουν στο φιλτράρισμα της πληροφορίας που προβάλλεται στους χρήστες, προτείνοντας μόνο τα στοιχεία που είναι πιθανό να τους ενδιαφέρουν. [41].

Ένα σημαντικό **μειονέκτημα** αυτού του αλγορίθμου είναι η έλλειψη αρχικής πληροφορίας, που οδηγεί σε αναξιόπιστα αποτελέσματα. Πιο συγκεκριμένα, ακόμη και οι πιο ενεργοί χρήστες ενός συστήματος συνήθως αξιολογούν μόνο ένα μικρό υποσύνολο των αντικειμένων που υπάρχουν στο σύστημα, με αποτέλεσμα οι πληροφορίες που θα συγκεντρωθούν πιθανότατα να είναι ανεπαρκείς για να επιτευχθούν αξιόπιστα αποτελέσματα.

Σημαντικά **πλεονεκτήματα** αυτού του αλγορίθμου είναι τα εξής [36]:

Δεν απαιτείται κατανάλωση πόρων του συστήματος για τη συγκέντρωση πληροφοριών σχετικά με τα αντικείμενα. Απουσιάζει η ανάγκη για λεπτομερή γνώση του αντικειμένου, καθώς το σύστημα βασίζεται αποκλειστικά στις αξιολογήσεις που δίνουν οι χρήστες για το συγκεκριμένο προϊόν, προκειμένου να προσφέρει σωστές συστάσεις.

Επιπλέον, λόγω της συνεχούς αύξησης του αριθμού των διαθέσιμων συστάσεων για ένα αντικείμενο, η ποιότητα των συστάσεων αυξάνεται, επιτρέποντας στο σύστημα να γίνεται όλο και πιο αξιόπιστο με την πάροδο του χρόνου.

Όσον αφορά γνωστά συστήματα συστάσεων, η Amazon χρησιμοποιεί αλγορίθμους συνεργατικού φιλτραρίσματος για να προτείνει προϊόντα στους χρήστες. Επίσης, τα κοινωνικά δίκτυα Facebook και LinkedIn χρησιμοποιούν συνεργατικό φιλτράρισμα για να προτείνουν στους χρήστες νέους φίλους ή ομάδες.

5.3.2 Συστήματα Συστάσεων Βάσει Περιεχομένου (Content-Based Recommender Systems)

Αυτοί οι αλγόριθμοι δημιουργούν συστάσεις προϊόντων με βάση τις προτιμήσεις που ο χρήστης έχει εκδηλώσει σε προηγούμενες αλληλεπιδράσεις του με το σύστημα. Συνεπώς, του προτείνονται αντικείμενα παρόμοια με αυτά που έχει προτιμήσει στο παρελθόν. Για κάθε χρήστη δημιουργείται σταδιακά και διατηρείται ένα προφίλ, το οποίο βασίζεται στην ανάλυση των προϊόντων που έχει επισκεφτεί ή αξιολογήσει.

Στους αλγορίθμους αυτούς, οι έννοιες του προφίλ χρήστη και του αντικειμένου παίζουν κεντρικό ρόλο. Κάθε χρήστης διατηρεί ένα προφίλ με τις επιλογές που έχει κάνει και τα αντικείμενα που τον ενδιαφέρουν. Αποτέλεσμα αυτού είναι ο αλγόριθμος να προτείνει αντικείμενα που είναι

παρόμοια με τις προηγούμενες επιλογές του χρήστη.

Τα Συστήματα Συστάσεων που βασίζονται στο περιεχόμενο επιδιώκουν να συσχετίσουν τους χρήστες με αντικείμενα που είναι παρόμοια με αυτά που τους άρεσαν στο παρελθόν. Η ομοιότητα αυτή δεν βασίζεται απαραιτήτως σε συνδέσεις αξιολόγησης μεταξύ των χρηστών, αλλά στα χαρακτηριστικά των αντικειμένων που αρέσουν στον χρήστη.

Αντίθετα από τα συνεργατικά συστήματα που χρησιμοποιούν εκτενώς τις αξιολογήσεις άλλων χρηστών σε συνδυασμό με αυτές του χρήστη, τα συστήματα που βασίζονται στο περιεχόμενο επικεντρώνονται κυρίως στις αξιολογήσεις του χρήστη και στις ιδιότητες των αντικειμένων που του αρέσουν. Ως εκ τούτου, οι άλλοι χρήστες έχουν ελάχιστο, αν όχι καθόλου, ρόλο στα συστήματα σύστασης που βασίζονται στο περιεχόμενο.Η μεθοδολογία που βασίζεται στο περιεχόμενο χρησιμοποιεί μια διαφορετική πηγή δεδομένων για τη διαδικασία της σύστασης.

Στο πιο βασικό επίπεδο, τα Συστήματα Συστάσεων που βασίζονται στο περιεχόμενο εξαρτώνται από δύο πηγές δεδομένων [42]:

1. Η πρώτη πηγή δεδομένων περιλαμβάνει περιγραφές διάφορων αντικειμένων που βασίζονται στα χαρακτηριστικά τους σύμφωνα με το περιεχόμενο. Ένα παράδειγμα αυτής της αναπαράστασης μπορεί να είναι η γραπτή περιγραφή ενός αντικειμένου από τον κατασκευαστή.

2. Η δεύτερη πηγή δεδομένων περιλαμβάνει ένα προφίλ χρήστη που δημιουργείται μέσω των σχολίων του χρήστη για διάφορα αντικείμενα. Η ανατροφοδότηση του χρήστη μπορεί να είναι ρητή ή σιωπηρή. Η ρητή ανατροφοδότηση αντιστοιχεί σε αξιολογήσεις, ενώ η σιωπηρή ανατροφοδότηση αντιστοιχεί στις ενέργειες του χρήστη. Οι αξιολογήσεις συλλέγονται με παρόμοιο τρόπο με τα συνεργατικά συστήματα. Το προφίλ χρήστη συνδέει τα χαρακτηριστικά των διάφορων αντικειμένων με τα ενδιαφέροντα του χρήστη (αξιολογήσεις).

Σημαντικά **πλεονεκτήματα** αυτού του αλγορίθμου είναι ότι εξετάζουν αποκλειστικά τις προτιμήσεις του χρήστη και αγνοούν εντελώς την συμπεριφορά των υπόλοιπων χρηστών. Επιπλέον οι αλγόριθμοι αυτοί μπορούν να προτείνουν αντικείμενα για τα οποία δεν έχει γίνει καμία κριτική ή βαθμολόγηση.

Τα συστήματα που βασίζονται στο περιεχόμενο έχουν κάποιους περιορισμούς όπως [43]:

1.Υπάρχουν περιορισμοί που προκύπτουν από την ανάλυση του περιεχομένου. Τα συστήματα που λειτουργούν με βάση το περιεχόμενό τους αντιμετωπίζουν δυσκολίες στην αξιολόγηση του περιεχομένου ενός βίντεο ή μιας φωτογραφίας. Η αξιολόγησή τους γίνεται με τη χρήση αυτοματοποιημένων αλγορίθμων και το αποτέλεσμα δεν είναι πάντα αξιόπιστο. Επιπλέον, σε κάποια κείμενα που έχουν το ίδιο προφίλ, τα συστήματα συστάσεων με βάση το περιεχόμενο δυσκολεύονται να διακρίνουν ποιο κείμενο είναι αυτό που χρειάζεται ο χρήστης και ποιο όχι.

2.Επιπλέον, αυτά τα συστήματα επιλέγουν πληροφορίες που έχουν υψηλή συσχέτιση με το προφίλ του χρήστη. Έτσι, αν ένας χρήστης αναζητά κάτι που δεν ταιριάζει με το προφίλ που έχει δημιουργήσει, τα αποτελέσματα θα είναι πολύ περιορισμένα σε σύγκριση με άλλους χρήστες. Για να αντιμετωπιστεί αυτό το πρόβλημα, έχουν αναπτυχθεί αλγόριθμοι που προσθέτουν τυχαιότητα στις προτάσεις του συστήματος.

3. Τέλος δημιουργούνται πολλά προβλήματα με τους νέους χρήστες οι οποίοι δεν έχουν συγκεντρώσει πολλά στοιχεία στο προφίλ τους και έτσι τα αποτελέσματα που έχουν από τα συστήματα συστάσεων δεν μπορούν να είναι αξιόπιστα.

5.3.3 Συστήματα Συστάσεων Βασισμένα Στη Γνώση (Knowledge-Based Recommender Systems)

Τα συστήματα σε αυτή τη κατηγορία βασίζονται στη γνώση που έχουν σχετικά με τα χαρακτηριστικά των αντικειμένων ή των χρηστών και χρησιμοποιούν αυτήν τη γνώση για να παράγουν κατάλληλες συστάσεις. Η γνώση που διαθέτουν τα συστήματα αυτά χωρίζεται σε τρεις κατηγορίες και ανάλογα με το ποια κατηγορία γνώσης διαθέτει κάθε σύστημα, λαμβάνονται οι αντίστοιχες ενέργειες.Οι τρεις κατηγορίες είναι [37]:

•Γνώση για τα **αντικείμενα**: Το σύστημα πρέπει να γνωρίζει λεπτομέρειες για το κάθε αντικείμενο που περιέχει για να μπορέσει να κάνει τις απαραίτητες συγκρίσεις και να κάνει τις απαραίτητες συστάσεις ενός προϊόντος αλλά και όλα τα άλλα προϊόντα που ανήκουν στην ίδια κατηγορία.

• Γνώση για τους **χρήστες**: Το σύστημα ακόμα πρέπει να γνωρίζει λεπτομέρειες σχετικά με τον κάθε χρήστη που είναι μέλος του για να

καταφέρει να δημιουργήσει ένα ικανοποιητικό προφίλ που θα το βοηθήσει κατά την διάρκεια της δημιουργίας συστάσεων.

• Γνώση του **τρόπου** με τον οποίο καλύπτονται οι ανάγκες: Το σύστημα πρέπει να γνωρίζει ποια προϊόντα μπορούν να καλύψουν τις ανάγκες των χρηστών και να κάνει την αντιστοίχηση με τα διαθέσιμα προϊόντα που υπάρχουν.

Η διαδικασία απόκτησης γνώσης διαφέρει ανάλογα με την κατηγορία των συστημάτων. Όσον αφορά τη γνώση σχετικά με τα αντικείμενα, το σύστημα πρέπει να αντλήσει πληροφορίες από τις βάσεις δεδομένων που περιέχει. Στη συνέχεια, το σύστημα αποφασίζει εάν θα χρησιμοποιήσει τις αποκτηθείσες πληροφορίες ή όχι. Αυτός ο τρόπος απόκτησης γνώσης ονομάζεται έμμεσος. Ταυτόχρονα, υπάρχει και ο άμεσος τρόπος, όπου οι διαχειριστές του συστήματος εισάγουν δεδομένα που λαμβάνουν από τους παραγωγούς ή τους πωλητές των προϊόντων. Ο άμεσος τρόπος συνήθως χρησιμοποιείται σε συστήματα που προσφέρουν προϊόντα προς πώληση.Η παρακάτω εικόνα περιγράφει τον τρόπο με τον οποίο γίνονται οι συστάσεις στα συστήματα βασισμένα στην γνώση.

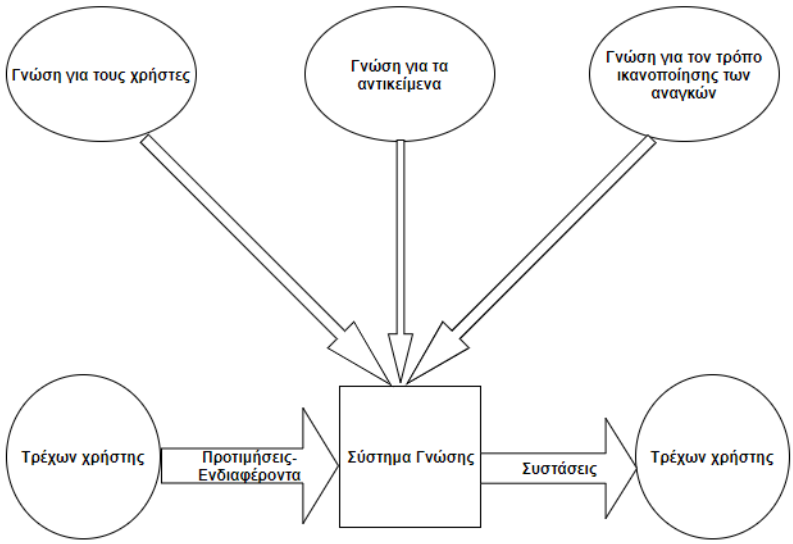

Figure 5.2: Περιγραφή Του Τρόπου Λειτουργίας Των Συστημάτων Βασισμένα Στην Γνώση

"Τροποποιήθηκε από την πηγή:
Μεθοδολογίες για τη δημιουργία συστημάτων σύστασης, Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής, Μπαλτζή Βασιλική

Τα βασισμένα στη γνώση συστήματα χρησιμοποιούνται συνήθως όταν το συνεργατικό φιλτράρισμα ή το φιλτράρισμα με βάση το περιεχόμενο δεν μπορούν να εφαρμοστούν.

Ένα σημαντικό πλεονέκτημα των αλγορίθμων που βασίζονται στη γνώση είναι ότι η απόδοσή τους δεν εξαρτάται από τον αριθμό των χρηστών ή τις αξιολογήσεις των χρηστών. Ωστόσο, σε αυτού του είδους αλγορίθμους είναι απαραίτητο τα δεδομένα να ανανεώνονται συνεχώς, έτσι ώστε το σύστημα να μπορεί να παράγει αξιόπιστες συστάσεις.

Τα συστήματα που βασίζονται στη γνώση έχουν κάποια μειονεκτήματα όπως:

Η κατασκευή ενός συστήματος συστάσεων βασισμένου στη γνώση απαιτεί μια προεργασία, η οποία είναι μακροχρόνια και απαιτεί επίμονη προσπάθεια. Πρέπει να δημιουργηθεί ένα σύστημα που μπορεί να κατανοήσει ένα αντικείμενο, μαζί με τα κύρια του χαρακτηριστικά και τις ανάγκες που ικανοποιεί.Επιπλέον, απαιτείται η δημιουργία μιας βάσης δεδομένων, όπου θα αποθηκεύονται όλα αυτά τα χαρακτηριστικά και θα είναι εύκολα και γρήγορα προσβάσιμα.Ωστόσο, σε περιπτώσεις όπου το σύστημα δεν περιλαμβάνει μεγάλη ποικιλία αντικειμένων, η πολυπλοκότητα των συστημάτων μειώνεται σημαντικά.

Τα συστήματα που βασίζονται στη γνώση χρησιμοποιούν την αρχική γνώση που έχουν δημιουργήσει για την παραγωγή συστάσεων. Ωστόσο, αυτά τα συστήματα δεν είναι σε θέση να αποκτήσουν νέα γνώση από τα δεδομένα που υπάρχουν στο σύστημα, εκτός αν η γνώση αυτή εισάγεται από τους διαχειριστές. Αυτό σημαίνει ότι απαιτείται ανθρώπινο δυναμικό για να ανανεώνεται η γνώση που διαθέτει το σύστημα και να παρέχει πιο ακριβείς συστάσεις..

Τέλος τα δεδομένα που διαθέτει το σύστημα θα πρέπει να είναι πολύ καλά δομημένα και οργανωμένα έτσι ώστε το σύστημα να είναι σε θέση να τα διαχειριστεί.

5.4 Συστήματα Συστάσεων με MLP

Τα συστήματα συστάσεων ανήκουν στην κατηγορία των μοντέλων μηχανικής μάθησης και έχουν ως στόχο να προτείνουν αντικείμενα ή προϊόντα

στους χρήστες με βάση την προηγούμενη συμπεριφορά, τις προτιμήσεις ή τις αλληλεπιδράσεις τους. Ένα είδος αρχιτεκτονικής νευρωνικού δικτύου που χρησιμοποιείται συχνά είναι το MLP (Multi-Layer Perceptron), το οποίο αποτελείται από πολλαπλά στρώματα διασυνδεδεμένων κόμβων. Τα στρώματα αυτά είναι ικανά να μάθουν σύνθετα πρότυπα από τα δεδομένα που εισέρχονται στο δίκτυο.

Η εφαρμογή του MLP στα συστήματα συστάσεων αποτελεί κοινή πρακτική, καθώς επιτρέπει ένα ευέλικτο και εκφραστικό μοντέλο που μπορεί να αποτυπώσει μη γραμμικές αλληλεπιδράσεις μεταξύ χρηστών και αντικειμένων.Μια κοινή προσέγγιση είναι η χρήση ενός MLP για την εκμάθηση μιας χαμηλής διάστασης αναπαράστασης (ενσωμάτωσης) τόσο των χρηστών όσο και των αντικειμένων, η οποία μπορεί να τροφοδοτηθεί σε ένα τελικό επίπεδο πρόβλεψης που εκτιμά την προτίμηση του χρήστη για κάθε αντικείμενο.

Συνολικά, η χρήση του MLP στα συστήματα συστάσεων είναι μια ισχυρή και ευέλικτη προσέγγιση που μπορεί να βελτιώσει την ακρίβεια και τη συνάφεια των συστάσεων.

5.4.1 Αλγόριθμος NeuMF (Neural Matrix Factorization)

Ο αλγόριθμος NeuMF (Neural Matrix Factorization) είναι μια τεχνική συνεργατικής φιλτράρισης σε συστήματα συστάσεων που συνδυάζει την παραγοντοποίηση πινάκων με νευρωνικά δίκτυα για τη βελτίωση της απόδοσης.Τα αρχικά NeuMF αντιπροσωπεύουν τον συνδυασμό αυτών των δύο τεχνικών για την ανάπτυξη του αλγορίθμου.

Η συνεργατική φιλτράριση αποτελεί μια τεχνική που χρησιμοποιείται σε συστήματα συστάσεων με σκοπό να προβλέψει τα ενδιαφέροντα ή τις προτιμήσεις ενός χρήστη.Αυτό επιτυγχάνεται μέσω της συλλογής πληροφοριών σχετικά με την προηγούμενη συμπεριφορά ή τις προτιμήσεις του συγκεκριμένου χρήστη και της σύγκρισής τους με τη συμπεριφορά άλλων χρηστών.Η Παραγοντοποίηση Πινάκων (Matrix Factorization - MF) αποτελεί μια δημοφιλή μέθοδο στη συνεργατική φιλτράριση που μαθαίνει τις κρυφές παραμέτρους που αναπαριστούν τους χρήστες και τα αντικείμενα.Αυτό επιτρέπει την παροχή εξατομικευμένων συστάσεων στους χρήστες.

1.Παραγοντοποίηση Πινάκων (Matrix Factorization): Η

παραγοντοποίηση πινάκων χρησιμοποιείται για την αναπαράσταση του πίνακα αλληλεπίδρασης μεταξύ χρήστη και αντικειμένου ως το γινόμενο δύο πινάκων, έναν πίνακα χαμηλής τάξης για τους χρήστες και έναν πίνακα χαμηλής τάξης για τα αντικείμενα. Ο πίνακας αλληλεπίδρασης συμβολίζεται ως R, ενώ οι πίνακες χαμηλής τάξης αναπαρίστανται ως P (για τους χρήστες) και Q (για τα αντικείμενα). Στόχος της παραγοντοποίησης πινάκων είναι να ελαχιστοποιήσει το σφάλμα πρόβλεψης μεταξύ του πίνακα R και του γινομένου των πινάκων P και Q.

2.Νευρωνικά Δίκτυα: Ο αλγόριθμος NeuMF χρησιμοποιεί νευρωνικά δίκτυα για να μοντελοποιήσει τις μη γραμμικές σχέσεις και τις πολύπλοκες αλληλεπιδράσεις μεταξύ των χρηστών και των αντικειμένων. Οι ακριβείς μαθηματικές εξισώσεις που χρησιμοποιούνται για τα νευρωνικά δίκτυα περιγράφουν τα επίπεδα, τις συναρτήσεις ενεργοποίησης και τις συνδέσεις μεταξύ των νευρώνων.

Συνδυασμός Εξόδων: Οι έξοδοι του στοιχείου παραγοντοποίησης πινάκων και του στοιχείου νευρωνικού δικτύου συνδυάζονται χρησιμοποιώντας πολλαπλασιασμό ή συνένωση στοιχείων(concatenation).Η τελική έξοδος τροφοδοτείται σε μια συνάρτηση ενεργοποίησης σιγμοειδούς για να προβλέψει τις προτιμήσεις του χρήστη για ένα συγκεκριμένο αντικείμενο.Το μοντέλο εκπαιδεύεται χρησιμοποιώντας αλγορίθμους ανατροφοδότησης προς τα πίσω (backpropagation) και τεχνικές βελτιστοποίησης.

Το NeuMF αξιοποιεί τόσο τα σήματα συνεργατικής φιλτράρισης από την παραγοντοποίηση πινάκων όσο και τα εξατομικευμένα σήματα που αποκομίζονται από το νευρωνικό δίκτυο, προσφέροντας ένα πιο ακριβές και αποτελεσματικό σύστημα συστάσεων.Έχει αποδειχθεί ότι παρουσιάζει βελτιωμένη απόδοση σε σύγκριση με τις παραδοσιακές μεθόδους παραγοντοποίησης πινάκων, προσφέροντας αυξημένη ακρίβεια στις συστάσεις και εξατομίκευση.

5.4.2 Μαθηματική εξίσωση του NeuMF (Neural Matrix Factorization)

Η εξίσωση του NeuMF αναπαριστά την πρόβλεψη της αξιολόγησης για ένα συγκεκριμένο χρήστη-αντικείμενο συνδυάζοντας τις εξόδους του στοιχείου παραγοντοποίησης πίνακα (MF) και του στοιχείου νευρωνικού δικτύου (MLP).

Συμβολίζουμε την έξοδο του MF ως MF(user, item) και την έξοδο του MLP ως MLP(user, item) [44].

Η εξίσωση του NeuMF μπορεί να εκφραστεί ως εξής:

$$NeuMF(user, item) = \alpha * MF(user, item) + \beta * MLP(user, item)$$

όπου α και β είναι συντελεστές που ρυθμίζουν την επίδραση του κάθε στοιχείου στην τελική πρόβλεψη. Αυτοί οι συντελεστές μπορούν να ρυθμιστούν κατάλληλα κατά την εκπαίδευση του μοντέλου.

Ο πιο απλός τύπος για τον NeuMF (Neural Matrix Factorization) μπορεί να περιγραφεί με την εξής μαθηματική έκφραση:

Έστω ότι έχουμε ένα σύνολο αξιολογήσεων ή βαθμολογίες (ratings) για τους χρήστες και τα αντικείμενα (π.χ. ταινίες) του συστήματος συστάσεων.Ο NeuMF εκτιμά τη βαθμολογία που θα έδινε ο χρήστης σε ένα συγκεκριμένο αντικείμενο ως το γινόμενο δύο συνιστωσών:

$$NeuMF_Rating = GMF_Rating + MLP_Rating$$

όπου:

GMF_Rating είναι η εκτίμηση βαθμολογίας που προκύπτει από το μοντέλο Matrix Factorization (GMF, Generalized Matrix Factorization).

MLP_Rating είναι η εκτίμηση βαθμολογίας που προκύπτει από το πολυεπίπεδο νευρωνικό δίκτυο (MLP, Multi-Layer Perceptron).

Ο GMF_Rating εκτιμάται από τον GMF ως το γινόμενο δύο παραγοντοποιημένων πινάκων:

GMF_Rating = Φ(GMF_User_Embedding $\odot$ GMF_Item_Embedding)

όπου:

• Φ είναι μια συνάρτηση ενεργοποίησης (π.χ. σιγμοειδής συνάρτηση ή γραμμική συνάρτηση).

• $\odot$ είναι ο στοιχειοθετικός πολλαπλασιασμός (element-wise multiplication).

• GMF_User_Embedding είναι η παράγοντοποίηση του χρήστη (user) σε έναν πίνακα.

• GMF_Item_Embedding είναι η παράγοντοποίηση του αντικειμένου (item) σε έναν πίνακα.

• Ο MLP_Rating εκτιμάται από το MLP ως η έξοδος του πολυεπίπεδου νευρωνικού δικτύου μετά από εφαρμογή των επιπέδων που έχουν οριστεί.

Συνολικά, ο τύπος του NeuMF αντιπροσωπεύει τη συνδυασμένη πρόβλεψη βαθμολογίας από το GMF και το MLP. Η τελική εκτίμηση NeuMF_Rating μπορεί να χρησιμοποιηθεί για να προβλέψει τις βαθμολογίες που θα έδινε ο χρήστης σε αντικείμενα που δεν έχει αξιολογήσει ακόμα.

5.5 Συστήματα Συστάσεων με RNN

Τα RNN είναι κατάλληλα για την επεξεργασία δεδομένων με διαδοχική σειρά.Ως αποτέλεσμα, αποτελούν τη βέλτιστη επιλογή όταν αναλύουμε τη χρονική δυναμική των αλληλεπιδράσεων και των συνεχόμενων μοτίβων στη συμπεριφορά των χρηστών.Επιπλέον, τα RNN είναι κατάλληλα για την επεξεργασία διαδοχικών σημάτων, όπως κείμενο, ήχος και άλλα.

Τα συστήματα συστάσεων που βασίζονται σε RNN μπορούν να κατηγοριοποιηθούν σε [45]:

Συστήματα βάσει συνεδρίας χωρίς αναγνωριστικό χρήστη: Στον πραγματικό κόσμο, η ταυτοποίηση του χρήστη δεν είναι πάντα απαραίτητη.Ωστόσο, τα συστήματα συστάσεων μπορούν να αποκτήσουν πληροφορίες σχετικά με τις προτιμήσεις του χρήστη χρησιμοποιώντας cookies.Για παράδειγμα, μπορούμε να ενσωματώσουμε έναν αλγόριθμο συστάσεων που λειτουργεί βάσει συνεδρίας, χρησιμοποιώντας κυρίως την τεχνική των Αναδραστικών Νευρωνικών Δικτύων (RNNs) για να προβλέψει τι θα αγοράσει ο χρήστης στο μέλλον, βασιζόμενος στις προηγούμενες αγορές του.

Συστήματα βάσει συνεδρίας με αναγνωριστικό χρήστη:Στα συστήματα διαδοχικών συστάσεων, όπου ο χρήστης αναγνωρίζεται από το

σύστημα, εφαρμόζονται διάφορες προσεγγίσεις. Για παράδειγμα, αναπτύσσεται ένα διαδραστικό σύστημα συστάσεων για την ανάκτηση δύο βασικών στόχων: την ερώτηση (ask) και την αντίδραση (react). Σε αυτό το σύστημα, τα RNN χρησιμοποιούνται για να προβλέψουν όχι μόνο τις απαντήσεις, αλλά και τις ερωτήσεις, λαμβάνοντας υπόψη την πρόσφατη δράση του χρήστη. Ένα άλλο παράδειγμα είναι ένα πολυ-εργαλειοποιημένο πλαίσιο μάθησης για την πρόβλεψη των χρόνων επιστροφής των χρηστών και την ταυτόχρονη σύσταση αντικειμένων.Σε αυτό το πλαίσιο, χρησιμοποιείται ένα μοντέλο ανάλυσης σε συνδυασμό με LSTM για τον υπολογισμό των χρόνων επιστροφής των πελατών.Επιπλέον, η πρόταση αντικειμένων βασίζεται στις προηγούμενες δραστηριότητες της συνεδρίας ενός χρήστη και χρησιμοποιεί επίσης LSTM.

Συνολικά, η χρήση των RNNs για την κατασκευή συστημάτων συστάσεων μπορεί να αποτελέσει μια αποτελεσματική προσέγγιση για τη μοντελοποίηση της συμπεριφοράς των χρηστών με την πάροδο του χρόνου και τη δημιουργία εξατομικευμένων συστάσεων με βάση αυτή τη συμπεριφορά. Ωστόσο, όπως συμβαίνει με κάθε μοντέλο μηχανικής μάθησης, είναι σημαντικό να αξιολογείται προσεκτικά και να συντονίζεται το μοντέλο ώστε να διασφαλίζεται ότι είναι ακριβές και αποτελεσματικό.

5.5.1 Αλγόριθμος GRU4Rec (Gated Recurrent Unit for Recommender Systems)

Ο αλγόριθμος GRU4Rec είναι μια παραλλαγή του αλγορίθμου Gated Recurrent Unit (GRU), ο οποίος χρησιμοποιείται συχνά για την ανάλυση ακολουθιών και τη σύσταση βασισμένη σε συνεδρίες.Στόχος του αλγορίθμου είναι να προτείνει αντικείμενα με βάση την τρέχουσα συνεδρία του χρήστη.Το μοντέλο GRU4Rec ενσωματώνει την αρχιτεκτονική του GRU για να καταγράφει διαδοχικά μοτίβα και να παρέχει εξατομικευμένες προτάσεις με βάση την προηγούμενη συμπεριφορά του χρήστη.

Στο μοντέλο GRU4Rec, κάθε περίοδος σύνδεσης του χρήστη αναπαρίσταται ως μια ακολουθία στοιχείων με τα οποία έχει αλληλεπιδράσει. Τα επίπεδα GRU χρησιμοποιούνται για να μοντελοποιήσουν τις χρονικές εξαρτήσεις μέσα στα δεδομένα της συνεδρίας. Οι μονάδες GRU περιέχουν μηχανισμούς πύλης που ρυθμίζουν τη ροή της πληροφορίας μέσα στο δίκτυο, επιτρέποντάς του να ανιχνεύει μακροπρόθεσμες εξαρτήσεις και να αντιμετωπίζει ακολουθίες εισόδου μεταβλητού μήκους.

Κατά τη διάρκεια της εκπαίδευσης, το μοντέλο GRU4Rec συνήθως βελτιστοποιείται χρησιμοποιώντας τον αλγόριθμο backpropagation through time (BPTT), ο οποίος επεκτείνει τον τυπικό αλγόριθμο backpropagation για να χειριστεί τις ακολουθίες. Ο στόχος είναι να ελαχιστοποιηθεί το σφάλμα πρόβλεψης μεταξύ του προβλεπόμενου στοιχείου και του επόμενου στοιχείου της σειράς.

Ο αλγόριθμος GRU4Rec έχει δείξει πολλά υποσχόμενη απόδοση σε εργασίες συστάσεων που βασίζονται σε περιόδους σύνδεσης, όπου ο στόχος είναι να προτείνει στοιχεία με βάση την τρέχουσα περίοδο σύνδεσης ενός χρήστη. Έχει εφαρμοστεί σε διάφορους τομείς, όπως το ηλεκτρονικό εμπόριο, οι ειδήσεις και οι προτάσεις μουσικής.

5.5.2 Μαθηματική εξίσωση του GRU4Rec (Gated Recurrent Unit for Recommender Systems)

Ο GRU4Rec (Gated Recurrent Unit for Recommender Systems) είναι ένα μοντέλο συστάσεων που χρησιμοποιείται για την πρόβλεψη ακολουθιών συστάσεων, όπως για παράδειγμα, σε ένα περιβάλλον ηλεκτρονικού εμπορίου όπου οι χρήστες κάνουν αγορές κατά σειρά χρόνου [46].

Ο πιο απλός τύπος για τον GRU4Rec αφορά την ενημέρωση των κρυφών καταστάσεών του σε κάθε βήμα της ακολουθίας:

Αρχικοποίηση: h_0 = 0

Για κάθε χρονικό βήμα t και κάθε είσοδο (π.χ.αντικείμενο) x_t της ακολουθίας:

$$z\_t = \sigma\left(W\_z \cdot x\_t + U\_z \cdot h\_t - 1\right)$$

$$r\_t = \sigma\left(W\_r \cdot x\_t + U\_r \cdot h\_t - 1\right)$$

h_t = (1 - z_t) · h_t-1 + z_t · tanh(W_h · x_t + U_h · (r_t ⊙ h_t-1))

Στον παραπάνω τύπο:

h_t είναι το διάνυσμα κρυφής κατάστασης του μοντέλου στο χρονικό βήμα t.

x_t είναι το διάνυσμα εισόδου (π.χ. αντικείμενο) στο χρονικό βήμα t.

h_t-1 είναι το διάνυσμα κρυφής κατάστασης στο προηγούμενο χρονικό βήμα (t-1).

z_t είναι η πύλη ενημέρωσης (update gate) που ρυθμίζει πόσο από την προηγούμενη κατάσταση θα μεταφερθεί στο νέο βήμα.

r_t είναι η πύλη επαναφόρτισης (reset gate) που ρυθμίζει πόσο από την προηγούμενη κατάσταση θα ληφθεί υπόψιν για τον υπολογισμό της νέας κατάστασης.

σ είναι η σιγμοειδής συνάρτηση ενεργοποίησης (σιγμοειδής συνάρτηση στα W_z, W_r, U_z, U_r).

tanh είναι η υπερβολική συνάρτηση ενεργοποίησης.

W_z, W_r, W_h είναι οι πίνακες βαρών που σχετίζονται με την είσοδο x_t.

U_z, U_r, U_h είναι οι πίνακες βαρών που σχετίζονται με το προηγούμενο κρυφό διάνυσμα h_t-1.

$\odot$ είναι ο στοιχειοθετικός πολλαπλασιασμός.

Ο παραπάνω τύπος περιγράφει την λειτουργία ενός μόνο GRU επιπέδου. Συνήθως, χρησιμοποιούνται πολλά GRU επίπεδα για να δημιουργηθούν πιο περίπλοκα νευρωνικά δίκτυα, ανάλογα με την ακριβή αρχιτεκτονική του GRU4Rec που χρησιμοποιείται.

5.6 Συστήματα Συστάσεων με A2C

Ο Actor δέχεται την αναπαράσταση της κατάστασης ως είσοδο και εξάγει την καλύτερη δράση στην τρέχουσα κατάσταση.Ο Actor έχει δύο είδη ενεργειών: 1) να κάνει μια ερώτηση ή 2) να προτείνει μια λίστα στοιχείων.

Μια πιθανή αρχιτεκτονική για τον Actor είναι ένα feed-forward

νευρωνικό δίκτυο. Με βάση την έξοδο του δικτύου, μπορούμε να αποφασίσουμε αν θα υποβληθεί μια ερώτηση ή αν θα γίνει μια σύσταση λίστας αντικειμένων. Ωστόσο, ο αριθμός των πιθανών αντικειμένων είναι συνήθως πολύ μεγαλύτερος από τον αριθμό των πιθανών ερωτήσεων, γεγονός που καθιστά τον Actor προκατειλημμένο προς την σύσταση αντικειμένων, ακόμη και όταν η ερώτηση του χρήστη δεν είναι σαφής. Επιπλέον, ο μεγάλος διακριτός χώρος δράσης καθιστά την εκπαίδευση μη αποδοτική και αναποτελεσματική για τα μοντέλα βασισμένα σε ενισχυτική μάθηση. Συνεπώς, ο Actor πρέπει να εξερευνήσει έναν μεγάλο διακριτό χώρο δράσης για να επιλέξει τα αντικείμενα-στόχους ή τις κατάλληλες ερωτήσεις που θα οδηγήσουν σε θετική ανταμοιβή, κάτι που αυξάνει την χρονική πολυπλοκότητα της λήψης αποφάσεων καθώς ο χώρος δράσης μεγαλώνει. Για να αντιμετωπιστούν αυτές οι προκλήσεις και να επιτευχθεί υψηλή αποτελεσματικότητα στη συνομιλία, ο δράστης βασίζεται σε ένα ιεραρχικό δέντρο ομαδοποίησης των αντικειμένων και των ερωτήσεων [14].

Επιβράβευση: Όταν ο Actor αναλαμβάνει δράση στο χρονικό βήμα t, θα λαμβάνει μια ανταμοιβή που δηλώνει πόσο καλή είναι η τρέχουσα δράση.Η τρέχουσα ενέργεια μπορεί να είναι η σύσταση ενός αντικειμένου ή η υποβολή μιας ερώτησης. Και στις δύο περιπτώσεις, η ενέργεια περιλαμβάνει ένα μονοπάτι από τη ρίζα του δέντρου προς τον κόμβο- φύλλο.

Ανταμοιβή για την υποβολή ερώτησης: Διαισθητικά, θέλουμε να κάνουμε ερωτήσεις που αυξάνουν την κατάταξη του αντικειμένου-στόχου έναντι άλλων αντικειμένων.Ωστόσο, στον μεγάλο χώρο αντικειμένων, αυτή η συνάρτηση ανταμοιβής δεν μπορεί να λειτουργήσει πολύ καλά και χρειάζεται πολύς χρόνος για να βελτιωθεί η κατάταξη του αντικειμένου-στόχου.Επιπλέον, ορισμένες ερωτήσεις μπορούν να βοηθήσουν στην εύρεση των υπο-ομάδες του αντικειμένου-στόχου και αυτό είναι ένα σημείο-κλειδί.Για παράδειγμα,ας υποθέσουμε ότι το ερώτημα είναι "φορτιστής για κινητά". Σε αυτή την περίπτωση, δύο πιθανές υπο-ομάδες μπορεί να είναι "ασύρματος" και "ενσύρματος".Επομένως, ο δράστης μπορεί να θέσει μια ερώτηση που σχετίζεται με αυτές τις δύο υπο-ομάδες για να διευκολύνει και να βρει το αντικείμενο-στόχο.

Για την αξιολόγηση της δράσης (π.χ. σύσταση ενός στοιχείου ή ερώτηση μιας ερώτησης) στην τρέχουσα κατάσταση σχεδιάζουμε έναν κριτικό που παίρνει τη κατάσταση και την αναληφθείσα δράση και μαθαίνει μια συνάρτηση δράσης-αξίας Q(s,a).Σύμφωνα με την Q(s,a), ο Actor ενημερώνει τις παραμέτρους του προκειμένου να παράγει πιο πολύτιμες ενέργειες στη συνομιλία.Η συνάρτηση δράσης-αξίας Q(s,a) είναι συνήθως μη γραμμική,

δεδομένου ότι οι χώροι δράσης και καταστάσεων είναι τεράστιοι.

5.6.1 Διαδικασία εκπαίδευσης και δοκιμής

Διαδικασία εκπαίδευσης κριτών(critic):

Η διαδικασία εκπαίδευσης του Critic στον αλγόριθμο A2C (Advantage Actor-Critic) στην ενισχυτική μάθηση περιλαμβάνει τα εξής βήματα:

Ορισμός του περιβάλλοντος (Environment): Αρχικά, πρέπει να ορίσουμε το περιβάλλον στο οποίο ο κριτής (Critic) θα αλληλεπιδρά με τον πράκτορα (Agent).Το περιβάλλον αναπαριστά το πρόβλημα που θέλουμε να λύσουμε, όπως για παράδειγμα το σύστημα συστάσεων, και παρέχει καταστάσεις και ανταμοιβές στον πράκτορα.

Αρχικοποίηση των παραμέτρων του Critic: Δημιουργούμε το νευρωνικό δίκτυο του Critic και αρχικοποιούμε τις παραμέτρους του με κάποια στρατηγική,όπως τυχαία αρχικοποίηση ή χρήση προ-εκπαιδευμένων παραμέτρων από άλλο μοντέλο.

Δημιουργία του συνόλου δεδομένων: Ο Critic χρειάζεται ένα σύνολο δεδομένων για την εκπαίδευσή του. Αυτό συνήθως προέρχεται από το περιβάλλον, δηλαδή τις καταστάσεις και τις ανταμοιβές που λαμβάνει ο πράκτορας κατά τη διάρκεια των αλληλεπιδράσεών του με το περιβάλλον.

Εκπαίδευση του Critic: Κατά τη διαδικασία εκπαίδευσης, παρουσιάζουμε το σύνολο δεδομένων στον Critic και τον καθοδηγούμε να εκτιμήσει τις πλεονεκτήματα (advantages) για κάθε κατάσταση και ενέργεια. Τα πλεονεκτήματα μπορούν να υπολογιστούν χρησιμοποιώντας μια κατάλληλη συνάρτηση που συγκρίνει την αναμενόμενη ανταμοιβή από μια κατάσταση με την εκτιμώμενη αξία της κατάστασης από τον Critic.

Βελτιστοποίηση του Critic: Οι παράμετροι του Critic ενημερώνονται χρησιμοποιώντας κάποιον αλγόριθμο βελτιστοποίησης, όπως το SGD (Stochastic Gradient Descent) ή κάποια παραλλαγή του, προκειμένου να ελαχιστοποιηθεί το σφάλμα εκτίμησης των πλεονεκτημάτων.

Επανάληψη της διαδικασίας: Τα παραπάνω βήματα επαναλαμβάνονται για πολλούς επαναληπτικούς κύκλους εκπαίδευσης, προκειμένου να βελτιστοποιηθούν οι παράμετροι του Critic και να βελτιωθεί η

απόδοσή του.

Διαδικασία εκπαίδευσης ηθοποιών(actor)

Η εκπαίδευση ηθοποιών (actors) στα Συστήματα Συστάσεων με τη χρήση του αλγορίθμου Advantage Actor-Critic (A2C) ακολουθεί μια παρόμοια διαδικασία με το στάδιο της εκπαίδευσης στο A2C που αναφέρθηκε προηγουμένως. Ωστόσο, υπάρχουν ορισμένες προσαρμογές και σημαντικές αλλαγές που αντικατοπτρίζουν το περιβάλλον των Συστημάτων Συστάσεων.

Οι βασικές φάσεις της διαδικασίας εκπαίδευσης ηθοποιών στα Συστήματα Συστάσεων με A2C περιλαμβάνουν τα εξής:

Περιβάλλον συστάσεων: Καθορίζεται το περιβάλλον στο οποίο θα λειτουργεί ο ηθοποιός συστάσεων. Αυτό μπορεί να είναι ένα σύστημα συστάσεων που προσφέρει συστάσεις σε χρήστες, όπως για παράδειγμα ένα ηλεκτρονικό κατάστημα ή μια πλατφόρμα streaming.

Περιγραφή καταστάσεων: Οι καταστάσεις του περιβάλλοντος συστάσεων περιγράφονται με βάση τα διαθέσιμα δεδομένα και τις παραμέτρους του περιβάλλοντος. Αυτές οι καταστάσεις μπορεί να περιλαμβάνουν πληροφορίες για τους χρήστες, τα προϊόντα ή τις υπηρεσίες που παρέχονται και τις προηγούμενες αλληλεπιδράσεις.

Καθορισμός πολιτικής: Ο ηθοποιός (actor) αναπτύσσει μια πολιτική που καθορίζει πώς θα επιλέγονται οι συστάσεις. Αυτή η πολιτική μπορεί να βασίζεται σε μηχανισμούς ενίσχυσης (reinforcement), μηχανικής μάθησης ή στατιστικών αλγορίθμων.

Εκπαίδευση με A2C: Ο ηθοποιός εκπαιδεύεται χρησιμοποιώντας τον αλγόριθμο Advantage Actor-Critic (A2C). Αυτό συμπεριλαμβάνει την αλληλεπίδραση του ηθοποιού με το περιβάλλον, την εκτίμηση των αποδόσεων (rewards) και την ανανέωση των παραμέτρων του ηθοποιού με βάση την κριτική των αξιολογητών (critics).

Επανάληψη και βελτίωση: Η διαδικασία της εκπαίδευσης επαναλαμβάνεται για πολλούς γύρους με σκοπό τη σταδιακή βελτίωση της απόδοσης του ηθοποιού. Οι παράμετροι του ηθοποιού ενημερώνονται κατάλληλα κατά τη διάρκεια αυτής της διαδικασίας βελτίωσης.

Τα παραπάνω βήματα συνθέτουν τη διαδικασία εκπαίδευσης ηθοποιών (actors) στα Συστήματα Συστάσεων με A2C. Η συνεχής επανάληψη και

βελτίωση του ηθοποιού οδηγεί σε μια πιο αποτελεσματική πολιτική συστάσεων, που μπορεί να προσαρμόζεται και να βελτιώνεται στις ανάγκες και τις προτιμήσεις των χρηστών.

Η διαδικασία δοκιμής

Κατά τη διάρκεια της δοκιμής, χρησιμοποιούμε απλώς τον ηθοποιό (Actor) θ_π ως το σύστημα συνομιλιακών συστάσεων. Κατά την είσοδο του ηθοποιού κατά τη δοκιμή, δίνουμε ένα αρχικό αίτημα. Σε κάθε γύρο της συνομιλίας, ο ηθοποιός μπορεί να θέσει μια ερώτηση ή να συστήσει έναν κατάλογο αντικειμένων. Εάν ο ηθοποιός αποφασίσει να συστήσει έναν κατάλογο αντικειμένων, επιλέγουμε τα 100 κορυφαία αντικείμενα και τα προσθέτουμε στην κατάταξή μας για την εν λόγω συνεδρία. Στους επόμενους γύρους, εάν ο ηθοποιός συστήσει ένα άλλο αντικείμενο, το προσθέτουμε σε αυτήν την κατάταξη. Εάν ο ηθοποιός επιλέξει να θέσει μια ερώτηση, επιστρέφουμε την απάντηση σε αυτήν την ερώτηση, εφόσον η απάντηση υπάρχει στην περιγραφή του αντικειμένου. Η συνομιλία θα τερματιστεί εάν ο ηθοποιός βρει το αντικείμενο-στόχο ή εάν ο αριθμός των γύρων της συνομιλίας υπερβαίνει τον μέγιστο επιτρεπτό αριθμό γύρων.

5.6.2 Αλγόριθμος REINFORCE

Ένας αλγόριθμος συστάσεων που συνδέεται συχνά με τον αλγόριθμο A2C είναι ο αλγόριθμος REINFORCE (ή πιο γενικά ο αλγόριθμος πολιτικής βελτιστοποίησης).

Ο αλγόριθμος REINFORCE αξιοποιεί τον A2C ως έναν τρόπο για τον υπολογισμό των πλεονεκτημάτων (advantages) και στη συνέχεια βελτιώνει την πολιτική του πράκτορα με βάση αυτά τα πλεονεκτήματα. Ο αλγόριθμος REINFORCE χρησιμοποιεί μια μοντελοποιημένη πολιτική (π.χ. πολυεπίπεδα νευρωνικά δίκτυα) για να επιλέξει ενέργειες, ενώ ο αλγόριθμος A2C παρέχει την εκτίμηση των πλεονεκτημάτων.

Ο αλγόριθμος REINFORCE είναι ένας αλγόριθμος ενισχυτικής μάθησης που χρησιμοποιείται για την εκπαίδευση πράκτορα να ανακαλύπτει τη βέλτιστη πολιτική (policy) για μια συγκεκριμένη εργασία. Ο αλγόριθμος REINFORCE είναι βασισμένος στη μέθοδο της πολιτικής ανανέωσης και ανήκει στην κατηγορία των αλγορίθμων που βασίζονται στην παραγώγιση.

Ο αλγόριθμος REINFORCE επιδιώκει να ελαχιστοποιήσει την

αντικειμενική συνάρτηση που αντιστοιχεί στην αναμενόμενη απόδοση της πολιτικής. Ο τρόπος υπολογισμού των ανταμοιβών, των γραμμικών πολλαπλασιαστών παραγώγων και η μέθοδος βελτιστοποίησης μπορούν να ποικίλουν ανάλογα με το συγκεκριμένο πρόβλημα που αντιμετωπίζετε.

5.6.3 Μαθηματική εξίσωση του Reinforce

Η πιο απλή μαθηματική εξίσωση για τον αλγόριθμο REINFORCE (Monte Carlo Policy Gradient) αφορά τον υπολογισμό του πλεονεκτήματος (Advantage) για μια επιλεγμένη ενέργεια και είναι η ακόλουθη [47]:

Advantage = (Σ όλων των μελλοντικών ανταμοιβών - Εκτιμώμενη Αξία της Κατάστασης πριν από την επιλογή της ενέργειας)

Σε αυτή την εξίσωση, το "Σ όλων των μελλοντικών ανταμοιβών" αναφέρεται στο άθροισμα όλων των μελλοντικών ανταμοιβών που λαμβάνει ο πράκτορας μετά την επιλογή της ενέργειας. Αυτό υπολογίζεται μετά την ολοκλήρωση της πολιτικής για ένα συγκεκριμένο πρόβλημα.

Επίσης, η "Εκτιμώμενη Αξία της Κατάστασης πριν από την επιλογή της ενέργειας" αναφέρεται στην πρόβλεψη του Critic για την αξία της κατάστασης στην οποία βρισκόταν ο πράκτορας πριν από την επιλογή της συγκεκριμένης ενέργειας.

Η εξίσωση του πλεονεκτήματος μας επιτρέπει να κατανοήσουμε πόσο καλή ή κακή ήταν η επιλεγμένη ενέργεια σε σχέση με τις προσδοκώμενες ανταμοιβές. Με βάση αυτό το πλεονέκτημα, ενημερώνουμε την πολιτική για να βελτιώσουμε τις μελλοντικές επιλογές του πράκτορα.

Advantage = (Σ όλων των μελλοντικών ανταμοιβών - Εκτιμώμενη Αξία της Κατάστασης πριν από την επιλογή της ενέργειας) * Διάνυσμα παραγόντων πολιτικής που αφορά τη συγκεκριμένη ενέργεια

Σε αυτήν την εξίσωση, οι βασικοί όροι παραμένουν οι ίδιοι όπως πριν, αλλά προστίθεται ένας πολλαπλασιαστικός όρος μεταξύ του πλεονεκτήματος και του διανύσματος παραγόντων πολιτικής που αφορά την συγκεκριμένη ενέργεια. Αυτός ο πολλαπλασιαστικός όρος λειτουργεί ως κλίση της πολιτικής ως προς τη συγκεκριμένη ενέργεια.

Ουσιαστικά, η εξίσωση αυτή μας δίνει το κίνητρο για την αύξηση ή

μείωση της πιθανότητας επιλογής συγκεκριμένων ενεργειών ανάλογα με το αν είναι θετικό ή αρνητικό το πλεονέκτημα. Αυτός είναι ο τρόπος με τον οποίο ο αλγόριθμος REINFORCE προσαρμόζει την πολιτική προς κατεύθυνση που βελτιώνει την επίδοση του πράκτορα με το πέρασμα του χρόνου.

5.7 Συστήματα Συστάσεων με DQN

Τα συστήματα συστάσεων που βασίζονται στη βαθιά μάθηση αναφέρονται σε μοντέλα και αλγόριθμους που χρησιμοποιούν τεχνικές βαθιάς μάθησης για την ανάπτυξη συστάσεων.Η βαθιά μάθηση είναι μια υποκατηγορία της μηχανικής μάθησης που αναφέρεται σε τεχνικές που χρησιμοποιούν τεχνητά νευρωνικά δίκτυα με πολλαπλά επίπεδα (βαθιά νευρωνικά δίκτυα) για την εξαγωγή υψηλού επιπέδου χαρακτηριστικών από τα δεδομένα εισόδου.

Στον τομέα των συστημάτων συστάσεων, η βαθιά μάθηση μπορεί να χρησιμοποιηθεί για να μοντελοποιήσει τον χρήστη και τα αντικείμενα με βάση ιδιότητες και συσχετίσεις που μπορούν να ανακαλυφθούν από τα δεδομένα. Με τη χρήση βαθιών νευρωνικών δικτύων, τα συστήματα συστάσεων μπορούν να αντιμετωπίσουν προβλήματα όπως η αντιμετώπιση του κρυμμένου χώρου διαστάσεων και η αντιμετώπιση μη γραμμικών συσχετίσεων μεταξύ των χαρακτηριστικών.

Οι διαδικασίες εκπαίδευσης συστημάτων συστάσεων με βάση τη βαθιά μάθηση συνήθως περιλαμβάνουν την προετοιμασία των δεδομένων, την ανάπτυξη του μοντέλου νευρωνικού δικτύου και την εκπαίδευση του μοντέλου χρησιμοποιώντας κατάλληλους αλγορίθμους βελτιστοποίησης, όπως ο αλγόριθμος A2C (Advantage Actor-Critic) [45] [48].

5.8 Συστήματα Συστάσεων με CNN

Το CNN είναι ένα νευρωνικό δίκτυο τροφοδότησης προς τα εμπρός που περιλαμβάνει επίπεδα συνέλιξης και διαδικασίες συγκέντρωσης.Είναι πιο ισχυρό όταν ασχολείται με μη δομημένα πολυμεσικά δεδομένα, όπως βίντεο, εικόνες ή κείμενο. Ειδικότερα, οι συστάσεις που βασίζονται στο CNN συνήθως έχει ως κύριο στόχο την εξαγωγή χαρακτηριστικών.Αυτή η εξαγωγή χαρακτηριστικών έχει προκύψει όπως παρακάτω:

Αναπαράσταση χαρακτηριστικών:

Εξαγωγή χαρακτηριστικών εικόνας: Ένα σύστημα συστάσεων CNN, όπου το μοντέλο τους εξετάζει τον αντίκτυπο των οπτικών χαρακτηριστικών στη σύσταση σημείων ενδιαφέροντος και παρουσιάζει ένα σύστημα σύστασης POI με οπτικό περιεχόμενο. Για την εξαγωγή χαρακτηριστικών εικόνας, το σύστημα αυτό χρησιμοποιεί CNN, το οποίο βασίζεται σε Probabilistic Matrix Factorization και διερευνά τις σχέσεις μεταξύ των οπτικών πληροφοριών και του λανθάνοντος παράγοντα θέσης/χρήστη.

Εξαγωγή χαρακτηριστικών κειμένου: Μια μεθοδολογία για τη σύσταση υλικού ηλεκτρονικής μάθησης. Χρησιμοποιούν CNN για την εξαγωγή χαρακτηριστικών στοιχείων από πληροφορίες κειμένου σε εκπαιδευτικό υλικό, όπως η εισαγωγή και η ουσία του υλικού.

Εξαγωγή χαρακτηριστικών ήχου και βίντεο: Προκειμένου να εξαχθούν χαρακτηριστικά από μουσικά σήματα,εισάγουμε τη χρήση CNNs που αντιμετωπίζουν την cold-start. Με τη χρήση αυτού του μοντέλου, είναι δυνατές πολλαπλές λειτουργίες χρονικών πλαισίων χάρη στους συνελικτικούς πυρήνες και τα στρώματα συγκέντρωσης.

Για να δημιουργηθεί ένα σύστημα συστάσεων με χρήση CNN, πρέπει να γίνουν διάφορα βασικά βήματα:

Συλλογή και προ-επεξεργασία δεδομένων: Το πρώτο βήμα είναι η συλλογή δεδομένων σχετικά με τους χρήστες και τα αντικείμενα στο σύστημα. Αυτό μπορεί να περιλαμβάνει πληροφορίες όπως περιγραφές αντικειμένων, κριτικές, εικόνες και αξιολογήσεις. Τα δεδομένα πρέπει επίσης να υποβληθούν σε προ επεξεργασία για την εξαγωγή σχετικών χαρακτηριστικών και τη μετατροπή τους σε μορφή που μπορεί να χρησιμοποιηθεί από το CNN.

Σχεδιασμός της αρχιτεκτονικής του νευρωνικού δικτύου: Το επόμενο βήμα είναι ο σχεδιασμός της αρχιτεκτονικής του νευρωνικού δικτύου. Η αρχιτεκτονική πρέπει να βελτιστοποιηθεί ώστε να εξάγει τα πιο σχετικά χαρακτηριστικά από τα δεδομένα εισόδου και να κάνει ακριβείς προβλέψεις. Η αρχιτεκτονική μπορεί να περιλαμβάνει πολλαπλά στρώματα συνελικτικών και συγκεντρωτικών στρωμάτων, καθώς και πλήρως συνδεδεμένα στρώματα.

Εκπαίδευση του νευρωνικού δικτύου: Αφού σχεδιαστεί η

αρχιτεκτονική, το επόμενο βήμα είναι η εκπαίδευση του νευρωνικού δικτύου. Αυτό περιλαμβάνει τη χρήση ενός μεγάλου συνόλου δεδομένων με αλληλεπιδράσεις χρήστη-αντικειμένου για τη βελτιστοποίηση των παραμέτρων του δικτύου. Κατά τη διάρκεια της εκπαίδευσης, το δίκτυο προσαρμόζεται ώστε να ελαχιστοποιεί τη διαφορά μεταξύ των προβλεπόμενων αξιολογήσεων και των πραγματικών αξιολογήσεων.

Αξιολόγηση του συστήματος συστάσεων: Αφού εκπαιδευτεί το νευρωνικό δίκτυο, πρέπει να αξιολογηθεί για να διαπιστωθεί πόσο καλά αποδίδει. Αυτό μπορεί να γίνει με τη σύγκριση των προβλεπόμενων αξιολογήσεων με τις πραγματικές αξιολογήσεις σε ένα σύνολο δοκιμών. Άλλες μετρικές όπως η ακρίβεια και η ανάκληση μπορούν επίσης να χρησιμοποιηθούν για την αξιολόγηση της απόδοσης του συστήματος συστάσεων.

Συνολικά, η χρήση CNN σε συστήματα συστάσεων μπορεί να οδηγήσει σε πιο ακριβείς και εξατομικευμένες συστάσεις για τους χρήστες. Ωστόσο, είναι σημαντικό να σχεδιάζεται προσεκτικά η αρχιτεκτονική και να βελτιστοποιούνται οι παράμετροι του δικτύου για να επιτυγχάνονται οι καλύτερες επιδόσεις. Είναι επίσης σημαντικό να ενημερώνεται και να βελτιώνεται συνεχώς το σύστημα καθώς διατίθενται νέα δεδομένα.

Τα συστήματα συστάσεων με χρήση Convolutional Neural Networks (CNN) είναι ένα είδος συστημάτων συστάσεων που χρησιμοποιούν CNN για να παράγουν συστάσεις για τους χρήστες με βάση τις προτιμήσεις τους ή τα προϊόντα που τους αρέσουν. Ο μαθηματικός τύπος που συνδέεται με αυτά τα συστήματα είναι ο τύπος της αντικειμενικής συνάρτησης (objective function) που χρησιμοποιείται για την εκπαίδευση του CNN.

Συγκεκριμένα, τα συστήματα συστάσεων με CNN συνήθως χρησιμοποιούν την αντικειμενική συνάρτηση απώλειας (loss function) για να εκπαιδεύσουν το μοντέλο. Η απώλεια είναι μια μετρική που μετρά τη διαφορά ανάμεσα στις πραγματικές συστάσεις (πραγματικές βαθμολογίες ή προτιμήσεις των χρηστών) και τις προβλέψεις του CNN μοντέλου. Ο στόχος κατά την εκπαίδευση είναι να ελαχιστοποιηθεί αυτή η απώλεια, ώστε το μοντέλο να παράγει όσο το δυνατόν πιο ακριβείς συστάσεις.

Μια απλή μορφή της αντικειμενικής συνάρτησης απώλειας που χρησιμοποιείται συχνά στα συστήματα συστάσεων είναι η τετραγωνική απώλεια (mean squared error - MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Σε αυτόν τον τύπο,το y_i αναπαριστά τις πραγματικές συστάσεις του χρήστη για το συγκεκριμένο αντικείμενο (πραγματική βαθμολογία) και το $\hat{y}_i$ αναπαριστά την πρόβλεψη του CNN μοντέλου για τις συστάσεις (προβλεπόμενη βαθμολογία).Το n αναφέρεται στον αριθμό των συστάσεων που χρησιμοποιούνται για την εκπαίδευση [49].

Κεφάλαιο 6

Επίλογος

Κατά τη διάρκεια αυτή της εργασίας παρουσιάσαμε τις βασικές αρχιτεκτονικές των Τεχνητών Νευρωνικών Δικτύων, τις διάφορες μεθόδους εκπαίδευσης τους καθώς και την εφαρμογή της μηχανικής και της ενισχυτικής μάθησης στα συστήματα συστάσεων. Η έρευνά μας ανέδειξε τη σημασία αυτών των τεχνικών στον καθορισμό προτεινόμενων αντικειμένων προς τους χρήστες, βελτιώνοντας σημαντικά την εμπειρία τους.

Στο 1ο κεφάλαιο, κάναμε μια εισαγωγή στη μηχανική μάθηση καθώς και στους τρόπους μάθησης καταλήγοντας στη σημασία της μηχανικής μάθησης ως ενός σημαντικού πεδίου στην εποχής μας.Η μηχανική μάθηση προσφέρει τεράστια δυνατότητα για τη λύση πολύπλοκων προβλημάτων και τη βελτίωση της απόδοσης συστημάτων. Είναι ένας τομέας που εξελίσσεται συνεχώς και ανοίγει νέες προοπτικές σε πολλούς τομείς, όπως η ρομποτική, η ιατρική, η αυτόνομη οδήγηση και πολλοί άλλοι.

Στο 2ο κεφάλαιο, αναδείξαμε τη σπουδαιότητα που προσφέρουν τα νευρωνικά δίκτυα πραγματοποιώντας μια σύντομη ιστορική αναδρομή.Επιπροσθέτως, δώσαμε ιδιαίτερη έμφαση στον τρόπο λειτουργίας του τεχνικού νευρώνα, στα χαρακτηριστικά του καθώς και στη σημασία των συναρτήσεων ενεργοποίησης τους, που αποτελούν ουσιώδες στοιχείο στη λειτουργία των νευρωνικών δικτύων.

Επίσης, αναλύσαμε τις διάφορες κατηγορίες των τεχνητών νευρωνικών δικτύων, καθεμία από τις οποίες έχει τις δικές της εφαρμογές και δυνατότητες. Αυτές οι κατηγορίες περιλαμβάνουν τα συνελικτικά νευρωνικά δίκτυα (CNN) που εξειδικεύονται στην ανάλυση εικόνων, τα αναδρομικά

νευρωνικά δίκτυα (RNN) που αντιμετωπίζουν ακολουθίες δεδομένων και τα πολυεπίπεδα νευρωνικά δίκτυα (MLP) που χρησιμοποιούνται για πολύπλοκες λειτουργίες μάθησης. Είναι κρίσιμο να κατανοούμε τις δυνατότητες και περιορισμούς κάθε κατηγορίας και να επιλέγουμε την κατάλληλη για κάθε εφαρμογή.

Η μηχανική μάθηση εξελίσσεται βασιζόμενη στην εξέλιξη αυτών των κατηγοριών, με τη σωστή εφαρμογή και συνεχή πρόοδο να οδηγούν σε εντυπωσιακές εφαρμογές σε πολλούς τομείς, όπως η αναγνώριση εικόνας, η αναγνώριση φωνής, η αυτόνομη οδήγηση και πολλά άλλα. Τέλος, αναλύσαμε το δίκτυο Perceptron τόσο ενός όσο και πολλών επιπέδων, και περιγράψαμε την γνωστή μέθοδο όπισθεν διάδοσης σφάλματος (backpropagation). Ο αλγόριθμος backpropagation αποτελεί τον κεντρικό πυρήνα της εκπαίδευσης των νευρωνικών δικτύων και είναι υπεύθυνος για τη διαδικασία της μάθησης.

Στο 3ο κεφάλαιο, επικεντρωθήκαμε στη σημασία της ενισχυτικής μάθησης στον τομέα της τεχνητής νοημοσύνης και αναλύσαμε τα βασικά στοιχεία της. Επιπροσθέτως, αναδείξαμε τη σημασία της μελέτης του αλγορίθμου Q-learning στον τομέα της ενισχυτικής μάθησης και της τεχνητής νοημοσύνης καθώς και τη μελέτη του αλγορίθμου Deep Q-learning.

Ο Q-learning αλγόριθμος αποτελεί έναν από τους πιο επιδραστικούς αλγορίθμους στον χώρο της αυτόνομης εκπαίδευσης αλγορίθμων και ρομποτικής.Με τον Q-learning, οι αυτόνομοι πράκτορες μπορούν να μάθουν πώς να επιτύχουν τους στόχους τους μέσα από δοκιμές και λάθη, χωρίς την ανάγκη της ανθρώπινης καθοδήγησης. Αυτό ανοίγει τον δρόμο για εφαρμογές στην αυτόνομη οδήγηση, την ρομποτική και πολλούς άλλους τομείς.

Τέλος, η ενισχυτική μάθηση αντιπροσωπεύει μια σημαντική προσέγγιση στον χώρο της μηχανικής μάθησης, όπου οι αλγόριθμοι μαθαίνουν να λύνουν προβλήματα μέσω αλληλεπιδράσεων με το περιβάλλον τους.Η ενισχυτική μάθηση είναι ένας τομέας της τεχνητής νοημοσύνης που συνεχώς εξελίσσεται και ανοίγει νέες προοπτικές για την ανάπτυξη ευφυών συστημάτων.

Στο 4ο κεφάλαιο, επιδιώκουμε να αναδείξουμε τη σπουδαιότητα των συστημάτων συστάσεων και την ποικιλία των βασικών μοντέλων συστημάτων συστάσεων που έχουν αναπτυχθεί. Τα κύρια μοντέλα συστημάτων συστάσεων, όπως αυτά που βασίζονται στη συνεργασία, το περιεχόμενο και τη γνώση, προσφέρουν διάφορες προσεγγίσεις για την παροχή συστάσεων. Κάθε ένα από αυτά τα μοντέλα διαθέτει τα δικά του πλεονεκτήματα και περιορισμούς και μπορεί να είναι κατάλληλο για διάφορες εφαρμογές. Στη

συνέχεια, παρουσιάσαμε μαθηματικούς τύπους για τα συστήματα συστάσεων ανά κατηγορία και αναφέραμε συγκεκριμένους αλγορίθμους συστάσεων για κάθε κατηγορία.

Συνοψίζοντας, αυτή η διπλωματική εργασία έχει αποκαλύψει τη σημασία της μηχανικής και της ενισχυτικής μάθησης στη βελτίωση των συστημάτων συστάσεων. Η συνεχής έρευνα σε αυτό τον τομέα μπορεί να οδηγήσει σε περαιτέρω καινοτομίες που θα βελτιώσουν την εμπειρία των χρηστών και θα έχουν ευρύτερες εφαρμογές στην κοινωνία.

Τέλος, θα ήθελα να ευχαριστήσω όλους όσους με υποστήριξαν κατά τη διάρκεια αυτής της διπλωματικής εργασίας και της ακαδημαϊκής μου πορείας. Οι συμβουλές, η υποστήριξη και η ενθάρρυνσή τους ήταν κρίσιμες για την επίτευξη αυτού του στόχου.

Βιβλιογραφία

[1] *"Τεχνητή Νοημοσύνη" - Μια εισαγωγική προσέγγιση*. el. Dec. 2016. URL: https://www.openbook.gr/techniti-noimosyni/ (visited on 02/17/2023).

[2] *ΚΕΦΑΛΑΙΟ 4 - Μηχανική Μάθηση*. URL: http://repfiles.kallipos.gr/html_books/93/04a-main.html (visited on 03/16/2023).

[3] Ροδάνθη Βιδάλη. «Αυτόματη Ταξινόμηση μελωδίας σε μουσικά είδη με τη χρήση τεχνητών νευρωνικών δικτύων». PhD thesis. Πανεπιστήμιο Πειραιώς - Τμήμα Πληροφορικής, 2011.

[4] ΕΛΕΝΗ-ΓΕΩΡΓΙΑ ΑΛΕΒΙΖΑΚΟΥ. «Η ΧΡΗΣΗ ΤΩΝ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΣΤΗΝ ΕΠΙΣΤΗΜΗ ΤΗΣ ΓΕΩΔΑΙΣΙΑΣ ΜΕ ΕΜΦΑΣΗ ΣΤΗΝ ΠΡΟΒΛΕΨΗ ΚΑΤΑΚΟΡΥΦΩΝ ΜΕΤΑΚΙΝΗΣΕΩΝ». PhD thesis. ΑΘΗΝΑ: ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ, 2012.

[5] *Νευρωνικό δίκτυο*. el. Page Version ID: 9497069. May 2022. URL: https://el.wikipedia.org/w/index.php?title=%CE%9D%CE%B5%CF%85%CF%81%CF%89%CE%BD%CE%B9%CE%BA%CF%8C_%CE%B4%CE%AF%CE%BA%CF%84%CF%85%CE%BF&oldid=9497069 (visited on 02/03/2023).

[6] Ηλίας Καταβάτης. «NEURAL NETWORKS AND LEARNING MACHINES». PhD thesis. ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ, 2019.

[7] Δήμητρα Σταθοπούλου. «Σύγκριση μεθόδων εκπαίδευσης τεχνητών νευρωνικών δικτύων». gr. In: (Dec. 2010). URL: https://hdl.handle.net/10889/4419 (visited on 02/17/2023).

[8] *linear transfer function definition - Google Search*. URL: https://www.google.com/search?q=linear+transfer+function+definition&sxsrf=AOaemvK-QPuwJRm8ViiJqZZeqfYDlmgDjg (visited on 02/03/2023).

[9] ΝΙΚΟΛΑΟΣ ΜΠΑΣΤΑΣ. «Επιλογή Χαρακτηριστικών και Ταξινόμηση Δεδομένων με την Χρήση Νευρωνικών Δικτύων RBF». PhD thesis. Θεσσαλονίκη: Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Α.Π.Θ., 2007.

[10] Χρήστος Κοντόπουλος. «Τεχνικές βαθιάς μηχανικής μάθησης και Συνελικτικά Νευρωνικά Δίκτυα για την ταξινόμηση υπερφασματικών δεδομένων». PhD thesis. Αθήνα: ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ, 2015.

[11] Δημήτριος Πήτας. «Νευρωνικά Δίκτυα και Εφαρμογές,Εφαρμογή Τεχνητών Νευρωνικών Δικτύων και Στατιστικών Μοντέλων στην Πρόβλεψη Χρονοσειρών». PhD thesis. Καρλόβασι: Σχολή Θετικών Επιστημών, ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ, 2020. URL: http://hdl.handle.net/11610/21484.

[12] Simon Haykin. *Neural Networks and Learning Machines*. Third Edition. Hamilton, Ontario, Canada, 2009.

[13] *Βασίλειος Χούτας*. June 2017. URL: https://www.slideshare.net/isselgroup/ss-77397656 (visited on 02/18/2023).

[14] Andrew Tch. *The mostly complete chart of Neural Networks, explained*. en. Aug. 2017. URL: https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464 (visited on 02/18/2023).

[15] *What are Recurrent Neural Networks? | IBM*. en-us. URL: https://www.ibm.com/topics/recurrent-neural-networks (visited on 02/18/2023).

[16] *Recurrent neural network*. en. Page Version ID: 1137152850. Feb. 2023. URL: https://en.wikipedia.org/w/index.php?title=Recurrent_neural_network&oldid=1137152850 (visited on 02/18/2023).

[17] Γεώργιος Μήτσιος. «Αναδρομικά Νευρωνικά Δίκτυα και αυτόματη παραγωγή Hashtag από Tweet του Twitter». PhD thesis. ΘΕΣΣΑΛΟΝΙΚΗ: ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ ΤΟΜΕΑΣ ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ, 2017.

[18] Στέφανος Π. Κανελλόπουλος. «Αναγνώριση Προσώπου με χρήση Συνελικτικών Νευρωνικών Δικτύων». ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ. ΑΘΗΝΑ: ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ, 2019.

[19] ΑΡΓΥΡΗΣ ΣΚΙΑΔΕΛΛΗΣ. «ΑΡΧΙΤΕΚΤΟΝΙΚΕΣ ΥΛΙΚΟΥ ΓΙΑ ΕΠΙΤΑΧΥΝΣΗ ΥΠΟΛΟΓΙΣΜΩΝ ΣΕ CONVOLUTIONS NEURAL NETWORKS». PhD thesis. Πάτρα: Πανεπιστήμιο Πατρών, 2021.

[20] Μαγδαληνή Κούγκουλα. «Ανίχνευση διαφορετικών αντικειμένων για αυτόνομες εφαρμογές οδήγησης». Μεταπτυχιακή Εργασία. ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ, 2022.

[21] *Δίκτυα Μακράς Βραχύχρονης Μνήμης*. el. Page Version ID: 9169785. Dec. 2021. URL: https://el.wikipedia.org/w/index.php?title=%CE%94%CE%AF%CE%BA%CF%84%CF%85%CE%B1_%CE%9C%CE%B1%CE%BA%CF%81%CE%AC%CF%82_%CE%92%CF%81%CE%B1%CF%87%CF%8D%CF%87%CF%81%CE%BF%CE%BD%CE%B7%CF%82_%CE%9C%CE%BD%CE%AE%CE%BC%CE%B7%CF%82&oldid=9169785 (visited on 02/18/2023).

[22] Βασίλειος Τσαλαβούτης Α. «Πειραματική διερεύνηση αλγορίθμων για βελτιστοποίηση της απόδοσης της πρόγνωσης χρονοσειρών με τη χρήση της μεθόδου των τεχνητών νευρωνικών δικτύων». ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ. Αθήνα: ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΤΟΜΕΑΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΕΡΕΥΝΑΣ, 2014.

[23] Ιωάννης Ιακωβίδης. «Εφαρμογές ενισχυτικής μάθησης στο πεδίο της δομημένης πρόβλεψης». PhD thesis. ΘΕΣΣΑΛΟΝΙΚΗ: ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ, ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ, 2017.

[24] *Ενισχυτική Μάθηση*. URL: https://www.intelligence.tuc.gr/~robots/ARCHIVE/2017w/projects/LAB51334503/rl.html (visited on 02/18/2023).

[25] *Ενισχυτική μάθηση*. el. Page Version ID: 6183699. Jan. 2017. URL: https://el.wikipedia.org/w/index.php?title=%CE%95%CE%BD%CE%B9%CF%83%CF%87%CF%85%CF%84%CE%B9%CE%BA%CE%AE_%CE%BC%CE%AC%CE%B8%CE%B7%CF%83%CE%B7&oldid=6183699 (visited on 02/18/2023).

[26] Ιωάννης Παρτάλας. «Μέθοδοι ενισχυτικής μάθησης σε συστήματα πρακτόρων». el. Διδακτορική Διατριβή. Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης (ΑΠΘ). Σχολή Θετικών Επιστημών. Τμήμα Πληροφορικής, 2009. DOI: 10.12681/eadd/19231. URL: http://hdl.handle.net/10442/hedi/19231 (visited on 02/18/2023).

[27] Ιωάννης Μπάτσιος. «Μηχανική και ενισχυτική μάθηση μέσω του αλγορίθμου Q-learning». el. bachelorThesis. 2021. URL: http://ir.lib.uth.gr/xmlui/handle/11615/55357 (visited on 02/21/2023).

[28] Ηλίας Αλκίδης. «Έλεγχος Της Οδικής Κυκλοφορίας Με Χρήση Αλγορίθμων Ενισχυτικής Μάθησης (reinforcement Learning)». Greek. In: (Oct. 2014). URL: http://artemis.cslab.ece.ntua.gr:8080/jspui/handle/123456789/12604 (visited on 02/18/2023).

[29] Shi Bichen. *Deep Reinforcement Learning in Recommender Systems with A3C*.

[30] Βασίλειος Στεργίου. «Συστήματα Συστάσεων με χρήση Βαθιάς Ενισχυτικής Μάθησης». PhD thesis. ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ: ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ, 2022.

[31] *Q-learning*. en. Page Version ID: 1135790537. Jan. 2023. URL: https://en.wikipedia.org/w/index.php?title=Q-learning&oldid=1135790537 (visited on 02/18/2023).

[32] Κωνσταντίνος Σκούρας. «Ενισχυτική Μάθηση και Αλγοριθμικές Συναλλαγές στο Χρηματιστήριο με την Τεχνική του Q-Learning». el. In: (Oct. 2019). URL: http://artemis.cslab.ece.ntua.gr:8080/jspui/handle/123456789/17405 (visited on 02/21/2023).

[33] Chathurangi Shyalika. *A Beginners Guide to Q-Learning*. en. July 2021. URL: https://towardsdatascience.com/a-beginners-guide-to-q-learning-c3e2a30a653c (visited on 02/21/2023).

[34] Ankit Choudhary. *Deep Q-Learning | An Introduction To Deep Reinforcement Learning*. en. Apr. 2019. URL: https://www.analyticsvidhya.com/blog/2019/04/introduction-deep-q-learning-python/ (visited on 02/25/2023).

[35] Y. Lin et al. «A Survey on Reinforcement Learning for Recommender Systems». In: *ArXiv* (Sept. 2021). URL: https://www.semanticscholar.org/paper/c2e26ea5abce9a96dee76923f791b40e0ba2a2fa (visited on 03/07/2023).

[36] Βασιλική Μπαλτζή. «ΓΡΑΦΟΘΕΩΡΗΤΙΚΕΣ ΜΕΘΟΔΟΛΟΓΙΕΣ ΓΙΑ ΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΣΥΣΤΗΜΑΤΩΝ ΣΥΣΤΑΣΗΣ». PhD thesis. Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής, 2015.

[37] R. Burke. «Knowledge-based recommender systems». In: 2000. URL: https://www.semanticscholar.org/paper/Knowledge-based-recommender-systems-Burke/dc133144d431fc3b75c8de27f6bb21da6eb5bc1b (visited on 02/25/2023).

[38] Christopher R. Aberger. «Recommender: An Analysis of Collaborative Filtering Techniques». In: (2014).

[39] Μάριος Μιχελής. «ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ: ΣΥΓΧΡΟΝΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ ΚΑΙ ΠΡΟΤΑΣΕΙΣ ΕΠΕΚΤΑΣΗΣ ΤΟΥΣ». PhD thesis. ΠΑΤΡΩΝ: ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ ΠΜΣ ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΥΠΟΛΟΓΙΣΤΩΝ ΕΦΑΡΜΟΓΩΝ ΚΑΙ ΘΕΜΕΛΙΩΣΕΩΝ ΤΗΣ ΕΠΙΣΤΗΜΗΣ ΤΩΝ ΥΠΟΛΟΓΙΣΤΩΝ, 2020.

[40] Pamela J. Ludford et al. «Think Different: Increasing Online Community Participation Using Uniqueness and Group Dissimilarity». In: (2004).

[41] Mark O'Connor et al. «PolyLens: A Recommender System for Groups of Users». In: *ECSCW 2001: Proceedings of the Seventh European Conference on Computer Supported Cooperative Work 16–20 September 2001, Bonn, Germany*. Ed. by Wolfgang Prinz et al. Dordrecht: Springer Netherlands, 2001, pp. 199–218. ISBN: 978-0-306-48019-5. DOI: 10.1007/0-306-48019-0_11. URL: https://doi.org/10.1007/0-306-48019-0_11.

[42] Charu C. Aggarwal. *Recommender Systems: The Textbook*. 1st ed. 2016. Cham: Springer International Publishing : Imprint: Springer, 2016. ISBN: 9783319296593.

[43] G. Adomavicius and A. Tuzhilin. «Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions». In: *IEEE Transactions on Knowledge and Data Engineering* 17.6 (2005), pp. 734–749. DOI: 10.1109/TKDE.2005.99.

[44] Chong Chen et al. «Efficient Neural Matrix Factorization without Sampling for Recommendation». In: *ACM Transactions on Information Systems* 38.2 (Jan. 2020), 14:1–14:28. ISSN: 1046-8188. DOI: 10.1145/3373807. URL: https://dl.acm.org/doi/10.1145/3373807 (visited on 08/10/2023).

[45] Shuai Zhang et al. «Deep Learning Based Recommender System: A Survey and New Perspectives». en. In: *ACM Computing Surveys* 52.1 (Jan. 2020), pp. 1–38. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/3285029. URL: https://dl.acm.org/doi/10.1145/3285029 (visited on 03/02/2023).

[46] Balázs Hidasi et al. *Session-based Recommendations with Recurrent Neural Networks*. arXiv:1511.06939 [cs]. Mar. 2016. URL: http://arxiv.org/abs/1511.06939 (visited on 08/10/2023).

[47] guest_blog. *REINFORCE Algorithm: Taking baby steps in reinforcement learning*. en. Nov. 2020. URL: https://www.analyticsvidhya.com/blog/2020/11/reinforce-algorithm-taking-baby-steps-in-reinforcement-learning/ (visited on 07/29/2023).

[48] Francesco Ricci, ed. *Recommender systems handbook*. OCLC: ocn373479846. New York: Springer, 2011. ISBN: 9780387858197 9780387858203.

[49] Kevin P. Murphy. *Machine learning: a probabilistic perspective*. eng. OCLC: 810414751. Cambridge, Mass.: MIT Press, 2012. ISBN: 9780262305242.