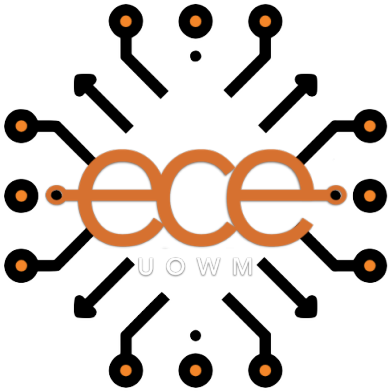


Πανεπιστήμιο Δυτικής Μακεδονίας

Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών

Αλγόριθμοι ενισχυτικής μάθησης για την εύρεση βέλτιστων στρατηγικών σε αγώνες αυτοκινήτου

Αλέξανδρος Δημητρίου (ΑΜ: 1317)

Επιβλέπων Καθηγητής: Νικόλαος Πλόσκας

Εργαστήριο Ευφυών Συστημάτων & Βελτιστοποίησης

20 Οκτωβρίου 2022

Περίληψη

Η ενισχυτική μάθηση είναι μία μορφή της μηχανικής μάθησης στην οποία ένας πράκτορας εκτελεί ενέργειες πάνω σε ένα περιβάλλον το οποίο εξερευνά και εκμεταλλεύεται τις γνώσεις που αποκτά με απώτερο σκοπό τη βελτιστοποίηση της επίδρασης των ενεργειών του. Η βελτιστοποίηση των ενεργειών του επιτυγχάνεται με την επαναληπτική εκτέλεσή τους στο περιβάλλον το οποίο κάθε φορά μας γνωστοποιεί για τα αποτελέσματα που προκάλεσε σε αυτό και κατά πόσο αποδοτική ήταν ενέργεια που εκτελέσθηκε. Στην παρούσα διπλωματική εργασία αφού αναφερθούν οι βασικές έννοιες της μηχανικής μάθησης και περιγράψουμε τη βασική δομή της ενισχυτικής μάθησης, αναφέρουμε πρακτικές μηχανικής μάθησης που έχουν αναπτύξει ή αναπτύσσουν καθημερινά ερευνητές του μηχανοκίνητου αθλητισμού. Πέρα από αυτά εξηγούμε τη δομή που έχει το πρόγραμμα πάνω στο οποίο εργαστήκαμε. Στη συνέχεια γίνεται περιγραφή της λειτουργίας του προγράμματος αυτού, όπου εδώ ο πράκτορας είναι ένας εικονικός μηχανικός στρατηγικής μίας ομάδας της Formula 1 με σκοπό τη δημιουργία ενός μοντέλου που θα κατατάξει τον οδηγό στην καλύτερη δυνατή θέση στο τέλος της προσομοίωσης. Συμπληρωματικά εξηγούμε τη δομή και τον τρόπο λειτουργίας της συνάρτησης ανταμοιβής που προτάθηκε στα πλαίσια της εργασίας. Έπειτα συγκρίνουμε τους μέσους όρους θέσεων που προκύπτουν για τον κάθε οδηγό, αφού γίνει προσομοίωση του κάθε αγώνα 10 φορές, στις οποίες ο πράκτορας εκμεταλλεύεται τα μοντέλα που έχουν δημιουργηθεί από τη συνάρτηση ανταμοιβής των δημιουργών και της συνάρτησης ανταμοιβής που προτάθηκε ως μέρος της εργασίας.

Λέξεις κλειδιά: Formula 1, ενισχυτική μάθηση, μηχανοκίνητος αθλητισμός, προσομοίωση, νευρωνικά δίκτυα

Abstract

Reinforcement learning is a part of machine learning where an agent performs actions in an environment which he explores and exploits the knowledge he acquires in order to maximize the effects of his actions. The maximization of the actions is achieved with recursive application of the actions in the environment which in every step returns the effects the actions had and how effective they were. In the present thesis after we mention the basic concepts of machine learning and describe the basic structure of reinforcement learning, we make a reference of machine learning applications which are applied or are in development from researchers of motor sport. Apart from these we explain the basic structure of program which we are studying. Then we describe how this program functions, where here the agent is a virtual strategy engineer of a Formula 1 team and aims to create a model which will put a driver in the best possible positions at the end of the simulation. In addition we explain the structure and how the suggested reward functions operates. Then, we compare the average positions of the drivers which result after the simulation of every race 10 times, in which the agent exploits the models which have been created by the reward function of the creators of the program and the suggested reward function.

Keywords: Formula 1, reinforcement learning, motor sport, simulation, neural networks

Δήλωση Πνευματικών Δικαιωμάτων

Δήλωση Πνευματικών Δικαιωμάτων Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν. 1599/1986 και τα άρθρα 2,4,6 παρ. 3 του Ν. 1256/1982, η παρούσα Διπλωματική Εργασία με τίτλο ''Αλγόριθμοι ενισχυτικής μάθησης για την εύρεση βέλτιστων στρατηγικών σε αγώνες αυτοκινήτου'' καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας και αναφέρονται ρητώς μέσα στο κείμενο που συνοδεύουν, και η οποία έχει εκπονηθεί στο Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών του Πανεπιστημίου Δυτικής Μακεδονίας, υπό την επίβλεψη του μέλους του Τμήματος κ. Νικολάου Πλόσκα αποτελεί αποκλειστικά προϊόν προσωπικής εργασίας και δεν προσβάλλει κάθε μορφής πνευματικά δικαιώματα τρίτων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή / και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και μόνο.

Υπογραφή Φοιτητή

Περιεχόμενα

Κατάλογος σχημάτων

Κεφάλαιο 1

Εισαγωγή

1.1 Ορισμός του προβλήματος

Όλοι μας γνωρίζουμε τις δυνατότητες τις οποίες έχει και μπορεί να προσφέρει η μηχανική μάθηση και πιο συγκεκριμένα η ενισχυτική μάθηση. Οι δυνατότητες που έχει στην ανάλυση και επεξεργασία μεγάλου όγκου δεδομένων φαίνεται στους τομείς στους οποίους έχει εφαρμοστεί. Όλοι οι τομείς που την έχουν εφαρμόσει έχουν σταθερό πλεονέκτημα απέναντι στους ανταγωνιστές τους και βρίσκονται πάντα ένα βήμα πιο μπροστά όσον αφορά την εξέλιξή τους.

Από τους τομείς οι οποίοι εφαρμόζουν μηχανική μάθηση δε θα μπορούσε να λείψει η Formula 1. Κατά τη διάρκεια ενός αγώνα της Formula 1 οι μηχανικοί που βρίσκονται εκτός του αγωνιστικού χώρου λαμβάνουν μεγάλο όγκο δεδομένων από το δικό τους μονοθέσιο, αλλά και των αντιπάλων τους. Έχοντας στη διάθεσή τους τόσα δεδομένα, υπάρχει μεγάλη πρόκληση για το τι θα λάβουν υπόψιν και τι όχι ώστε να επιλέξουν την κατάλληλη στρατηγική του οδηγού τους, στοχεύοντας στην καλύτερη δυνατή θέση στο τέλος του αγώνα.

1.2 Κίνητρα και στόχοι υλοποίησης

Έχοντας ως κίνητρο το μεγάλο ενδιαφέρον για τον μηχανοκίνητο αθλητισμό και για τις νέες πρακτικές της μηχανικής μάθησης δημιουργήθηκε η παρούσα εργασία στην οποία αφού γίνει κατανόηση ενός προγράμματος μηχανικής μάθησης πάνω στην Formula 1, γίνεται πρόταση μίας καινούργιας συνάρτησης ανταμοιβής η οποία συγκρίνεται με τη συνάρτηση ανταμοιβής των δημιουργών του προγράμματος ως προς το ποια από τα δύο μοντέλα που παράγονται χρησιμοποιώντας αυτές τις

συναρτήσεις θα κατατάξει τον οδηγό στην καλύτερη δυνατή θέση της τελικής κατάταξης του αγώνα έπειτα από αρκετές προσομοιώσεις των μοντέλων σε αυτόν.

1.3 Διάρθρωση κειμένου

Τα υπόλοιπα κεφάλαια οργανώνονται ως εξής: Στο δεύτερο κεφάλαιο παρουσιάζονται βασικά κομμάτια της μηχανικής μάθησης και γίνεται περιγραφή σημαντικών κομματιών για την εκπαίδευση ενός πράκτορα ενισχυτικής μάθησης. Στο τρίτο κεφάλαιο παρουσιάζονται οι διάφορες εφαρμογές που έχουν αναπτύξει και προτείνει ερευνητές του μηχανοκίνητου αθλητισμού. Στο τέταρτο κεφάλαιο γίνεται περιγραφή της δομής και λειτουργίας του προγράμματος πάνω στο οποίο εργαστήκαμε και της συνάρτησης ανταμοιβής που προτάθηκε. Στο πέμπτο κεφάλαιο γίνεται υπολογιστική μελέτη των αποτελεσμάτων τα οποία παρουσιάζονται σχηματικά για τον κάθε οδηγό ξεχωριστά και γίνεται σύγκριση για την αποδοτικότητα των μοντέλων που προκύπτουν από τις δύο συναρτήσεις ανταμοιβής. Στο έκτο και τελευταίο κεφάλαιο δίνεται μία προοπτική για την εξέλιξη και εφαρμογή τέτοιου είδους προγραμμάτων, όπως σε αυτό που εργαστήκαμε στην παρούσα εργασία, πάνω στον μηχανοκίνητο αθλητισμό.

Κεφάλαιο 2

Βιβλιογραφική Ανασκόπηση

2.1 Μηχανική μάθηση

Στην εποχή μας ο όγκος δεδομένων που υπάρχουν αποθηκευμένα σε διαφόρων τύπων υπολογιστών, από διάφορους τομείς της καθημερινής μας ζωής είναι τεράστιος. Πέρα από τους επιστήμονες όλοι μας έχουμε κατανοήσει ότι η εκμετάλλευση όλων αυτών των δεδομένων θα λύσει αρκετά προβλήματα της καθημερινότητάς μας. Αυτό τον ρόλο έρχεται να αναλάβει η μηχανική μάθηση. Ο όρος μηχανική μάθηση χρησιμοποιήθηκε για πρώτη το 1959 από τον Arthur Samuel και την όρισε ως πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί.

Η μηχανική μάθηση αναζητά να αναπτύξει υπολογιστικά συστήματα τα οποία αυτόματα βελτιώνουν την επίδοσή τους μέσω της εμπειρίας, μαθαίνοντας από τα λάθη τους. Η ιδέα της μηχανικής μάθησης αντιστοιχεί στην προσομοίωση της συμπεριφοράς των ανθρώπων. Δηλαδή προσπαθούμε να προσομοιώσουμε τη διαδικασία εκμάθησης και πως αντιδρά ο άνθρωπος σε κάθε ερέθισμα που λαμβάνει. Αυτό επιτυγχάνεται με τους αλγόριθμους της μηχανικής μάθησης οι οποίοι χρησιμοποιούν δεδομένα εισόδου ώστε να εκτελέσουν μία εργασία χωρίς να έχουν στην ουσία προγραμματιστεί για αυτή την εργασία, ώστε να παράξουν ένα συγκεκριμένο αποτέλεσμα. Οι αλγόριθμοι αυτοί αυτόματα αλλάζουν ή προσαρμόζονται με βάση την εμπειρία που αποκτούν σε κάθε πρόβλημα ώστε να γίνουν καλύτεροι στην εκτέλεση της εργασίας τους.

Η προσαρμογή και η βελτίωση αυτών των αλγορίθμων μπορεί να γίνει με διάφορους τρόπους. Μία λύση είναι τα δεδομένα εισόδου να είναι επιλεγμένα και

σταθμισμένα. Μπορούν να περιέχουν αριθμητικές μεταβλητές οι οποίες προσαρμόζονται μέσω της επαναληπτικής βελτιστοποίησης. Μπορεί να έχει ένα δίκτυο από πιθανά υπολογιστικά μονοπάτια τα οποία κατατάσσουν οι αλγόριθμοι για το βέλτιστο αποτέλεσμα. Τέλος μπορούν να δημιουργήσουν κατανομές πιθανοτήτων από τα δεδομένα εισόδου και να τα χρησιμοποιήσουν ώστε να προβλέψουν τα αποτελέσματα. Με βάση αυτών των μεθόδων προκύπτουν και οι διαφορετικοί τύποι μορφών μηχανικής μάθησης οι οποίοι είναι[1]:

- Επιβλεπόμενη μάθηση (Supervised Learning)
- Μη επιβλεπόμενη μάθηση (Unsupervised Learning)
- Ενισχυτική μάθηση (Reinforcement Learning)

2.1.1 Επιβλεπόμενη μάθηση

Η επιβλεπόμενη μάθηση είναι μία τεχνική στην οποία εισάγουμε επισημασμένα δεδομένα στο μοντέλο της μηχανικής μάθησης για να προβλέψουμε ένα σίγουρο αποτέλεσμα. Χρησιμοποιεί ένα σύνολο δεδομένων ώστε να μάθει στο μοντέλο πιο είναι το επιθυμητό αποτέλεσμα. Το σύνολο των δεδομένων περιλαμβάνει εισόδους και εξόδους που βοηθάνε στο μοντέλο να μάθει σε ποια είσοδο αντιστοιχεί το σωστό αποτέλεσμα. Έτσι δημιουργείται ένα μοντέλο το οποίο για κάθε καινούργια είσοδο δίνει το αποτέλεσμα που έχουμε προβλέψει κατά την εκπαίδευσή του. Αυτό γίνεται με τη χρήση μια συνάρτησης ανταμοιβής όπου κάθε φορά ο αλγόριθμος εξετάζει το αποτέλεσμά της ώστε να καταλήξει στο μικρότερο αποτέλεσμα λάθους μεταξύ εισόδου και εξόδου.

Η επιβλεπόμενη μάθηση μπορεί να χωριστεί σε δύο κατηγορίες επίλυσης προβλημάτων, της κατηγοριοποίησης που χρησιμοποιεί έναν αλγόριθμο ώστε να κατηγοριοποιήσει με ακρίβεια δεδομένα του εκάστοτε προβλήματος σε συγκεκριμένες κατηγορίες και η παλινδρόμηση όπου προσπαθεί να κατανοήσει τη σχέση μεταξύ των εξαρτώμενων και μη εξαρτώμενων μεταβλητών. Οι αλγόριθμοι που μπορούν να χρησιμοποιηθούν σε κάθε περίπτωση είναι η γραμμική παλινδρόμηση (Linear Regression), τα νευρωνικά δίκτυα (Neural Networks), οι μηχανές διανυσμάτων στήριξης (Support Vector Machines – SVMs), η μάθηση κατά Bayes (Bayesian Learning), τα δένδρα απόφασης (Decision Trees), ο k πλησιέστεροι γείτονες (k Nearest Neighbors –

kNN), η λογιστική παλινδρόμηση (Logistic Regression) και τα τυχαία δάση (Random Forests) [2].

2.1.2 Νευρωνικά δίκτυα

Τα νευρωνικά δίκτυα εξετάζουν δεδομένα μιμούμενα τη διασυνδεσημότητα που έχει ο ανθρώπινος εγκέφαλος μέσω στρωμάτων κόμβων. Οι κόμβοι αυτοί ονομάζονται νευρώνες και οι συνδέσεις μεταξύ τους ονομάζονται συνάψεις. Ο κάθε κόμβος αποτελείται απο εισόδους, βάρη, ένα όριο και μία έξοδο. Στην πιο απλή μορφή αποτελείται από δύο ξεχωριστά στρώματα νευρώνων τα οποία είναι είσοδος και η έξοδος. Οι νευρώνες του στρώματος της εξόδου λαμβάνουν σήματα από το στρώμα εισόδου αλλά το αντίθετο δεν μπορεί να συμβεί. Ένας κόμβος μπορεί να είναι συνδεμένος με πολλούς κόμβους από τους οποίους λαμβάνει δεδομένα και με πολλούς κόμβους στους οποίους στέλνει δεδομένα. Υπάρχουν και δίκτυα νευρώνων με πολλαπλά στρώματα όπου σε κάθε σύνδεση που φθάνει σε ένα κόμβο, αυτός αναθέτει έναν αριθμό οποίος ονομάζεται βάρος. Όταν το δίκτυο λειτουργεί ο κόμβος λαμβάνει διαφορετικά δεδομένα με διαφορετικά βάρη, οπότε για κάθε σύνδεση πολλαπλασιάζει τον αριθμό αυτό με καθορισμένο βάρος του και προσθέτει τον αριθμό με το αποτέλεσμα του πολλαπλασιασμού. Αν το αποτέλεσμα είναι πάνω από ένα συγκεκριμένο όριο που έχουμε θέσει τότε περνάει τα δεδομένα(τα βάρη) στους επόμενους κόμβους που βρίσκονται στο επόμενο στρώμα.

Όταν ένα δίκτυο νευρώνων εκπαιδεύεται τότε όλα τα βάρη και τα όρια που έχουμε θέσει παίρνουν τυχαίες τιμές. Τα δεδομένα για την εκπαίδευσή του προωθούνται στο στρώμα εισόδου όπου έπειτα γίνονται τα αθροίσματα και οι πολλαπλασιασμοί ώστε να καταλήξουμε να λάβουμε την έξοδο που θέλουμε στο στρώμα εξόδου. Καθόλη την εκπαίδευση τα βάρη και τα όρια αλλάζουν συνεχώς τιμές μέχρι τα δεδομένα εισόδου που έχουν ετικέτες με παρόμοια δεδομένα να παράγουν παρόμοιες εξόδους [3].

2.1.3 Μη επιβλεπόμενη μάθηση

Η μη επιβλεπόμενη μηχανική μάθηση χρησιμοποιεί αλγόριθμους για να αναλύει και να ομαδοποιεί δεδομένα τα οποία δεν είναι κατηγοριοποιημένα από πριν. Χωρίς να του παρέχεται καμία εμπειρία ο αλγόριθμος μάθησης πρέπει να βρει τη δομή

των δεδομένων εισόδου ώστε να τα κατατάξει και να έχουμε σωστή αναγνώριση των στοιχείων που θα εισέρχονται στον αλγόριθμο. Η μη επιβλεπόμενη μάθηση είναι παρόμοια με τον τρόπο με τον οποίο οι άνθρωποι μαθαίνουν από τις εμπειρίες τους όπου έχουν κατηγοριοποιήσει πολλά από τα αντικείμενα που βλέπουν καθημερινά και τα αναγνωρίζουν κάθε φορά που τα συναντούν ξανά. Αφού δουλεύει με δεδομένα εισόδου τα οποία δεν ανήκουν σε κάποια γνωστή για τον αλγόριθμο κατηγορία, είναι αρκετά χρήσιμη στην καθημερινή ζωή αφού δεν έχουμε πάντα δεδομένα για τα οποία γνωρίζουμε την έξοδός τους.

Ο αλγόριθμος μη επιβλεπόμενης μάθησης μπορεί να χωριστεί σε δύο τύπους προβλημάτων. Ένας είναι της ομαδοποίησης όπου συγκεντρώνει όλα τα στοιχεία με ομοιότητες σε ένα γκρουπ ξεχωριστά από άλλο γκρουπ δεδομένων τα οποία δεν έχουν καμία ομοιότητα με το προηγούμενο γκρουπ. Το δεύτερο είναι ο κανόνας συσχέτισης όπου χρησιμοποιείται για να βρει τις σχέσεις μεταξύ μεταβλητών σε μία μεγάλη βάση δεδομένων. Έχει τη δυνατότητα να καθορίσει ποια σετ δεδομένων εμφανίζονται μαζί. Αυτό είναι χρήσιμο σε στρατηγικές μάρκετινγκ όπου έχει τη δυνατότητα να συσχετίσει ότι ένα άτομο που αγοράζει ένα Χ προϊόν συχνά αγοράζει και το Ψ προϊόν. [4].

2.1.4 Ενισχυτική μάθηση

Η ενισχυτική μάθηση είναι ένα κομμάτι της μηχανικής μάθησης όπου στόχος είναι ένας πράκτορας να μάθει να συμπεριφέρεται σε ένα περιβάλλον στο οποίο λαμβάνει μόνο σήματα ανταμοιβής. Η ενισχυτική μάθηση δεν θα πρέπει να χαρακτηρίζεται από μία κλάση με μεθόδους μάθησης αλλά ως ένα πρόβλημα εκμάθησης. Ο στόχος του πράκτορα είναι να εκτελεί ενέργειες οι οποίες μεγιστοποιούν μακροπρόθεσμα τα σήματα ανταμοιβής. Ο πράκτορας μπορεί να διαλέξει μία ενέργεια σε κάθε κατάσταση και η αντίδραση που λαμβάνει από το περιβάλλον είναι η κατάσταση που έχει μετά από αυτή την ενέργεια. Ακόμα λαμβάνει την τιμή του σήματος της ανταμοιβής σε κάθε βήμα[5].

Τα κύρια μέρη ενός συστήματος ενισχυτικής μάθησης είναι ο πράκτορας, το περιβάλλον με το οποίο αντιδρά ο πράκτορας, την τακτική (policy) που ακολουθεί ο πράκτορας, τη συνάρτηση ανταμοιβής (reward) και την παρατήρηση (observation) του πράκτορα μέχρι να λάβει κάποια ενέργεια. Η συνάρτηση ανταμοιβής μας δείχνει

κατά πόσο καλή είναι η ενέργεια που έκανε ο πράκτορας δηλαδή πως επωφεληθήκαμε από αυτό. Ο στόχος του αλγόριθμου ενισχυτικής μάθησης είναι να βρεί μια τακτική η οποία θα μεγιστοποιεί την ανταμοιβή σε κάθε κατάσταση του συστήματος.

Μία αφηρημένη μορφή της συνάρτηση ανταμοιβής είναι η συνάρτηση τιμής όπου καταγράφει το πόσο καλή είναι μία κατάσταση. Η ανταμοιβή μας δείχνει το αποτέλεσμα που έχει μια ενέργεια σε μία συγκεκριμένη κατάσταση, όμως η συνάρτηση τιμής περιέχει αθροιστικά την ανταμοιβή από αυτή την κατάσταση και μετά(collect step).

Οι αλγόριθμοι της ενισχυτικής μάθησης χωρίζονται σε δύο κατηγορίες, αυτούς βασίζονται σε μοντέλα και αυτούς που δεν βασίζονται. Οι αλγόριθμοι που δεν βασίζονται σε μοντέλα είναι είτε βασισμένοι στην τιμή ή στην τακτική. Οι αλγόριθμοι που βασίζονται στην τιμή θεωρούν ότι η καλύτερη τιμή τακτικής (policy) είναι αποτέλεσμα υπολογισμού της συνάρτησης τιμής σε κάθε κατάσταση. Χρησιμοποιώντας επαναληπτικές συσχετίσεις ο αλγόριθμος αλληλοεπιδρά με το περιβάλλον ώστε να λάβει δεδομένα των καταστάσεων και τιμές της ανταμοιβής. Όταν θα έχει αρκετά δεδομένα μπορεί να υπολογίσει τη συνάρτηση τιμής. Αφού λάβει αυτή την τιμή, κάνει άπληστες ενέργειες ώστε να βρει τη βέλτιστη τακτική σε κάθε κατάσταση. Από την άλλη πλευρά οι αλγόριθμοι που βασίζονται στην τακτική υπολογίζουν την κατάλληλη τιμή της τακτικής χωρίς να χρειάζεται να μοντελοποιήσουν τη συνάρτηση τιμής. Παραμετροποιούν την τακτική απευθείας χρησιμοποιώντας μαθησιακά βάρη και μετατρέπουν το πρόβλημα εκμάθησης σε ένα πρόβλημα βελτιστοποίησης όπου στόχος είναι να αυξήσουμε τον ανώτατο βαθμό του μέσου όρου της συνάρτησης τιμής σε όλες τις καταστάσεις [6]. Τα κύρια μέρη της ενισχυτικής μάθησης που περιγράψαμε και η σύνδεσή τους φαίνεται στο Σχήμα 2.1.

Τις περισσότερες φορές δεν έχουμε γνώση του μοντέλου ενός προβλήματος και είναι απαραίτητο να διαδράσουμε με το περιβάλλον και να μάθουμε με δοκιμές και λάθη. Ο πράκτορας πρέπει να εξερευνήσει το περιβάλλον εκτελώντας ενέργειες και να λάβει τις κατάλληλες αλλαγές στο περιβάλλον και την ανταμοιβή. Ο πράκτορας λαμβάνει μόνο την πληροφορία της ανταμοιβής χωρίς να ξέρει αν είναι η σωστή ενέργεια που εκτέλεσε. Σε κάποιο σημείο των δοκιμών θα έχει μια τιμή της τακτικής με καλή επίδοση και θα προσπαθήσει να τη βελτιώσει. Ωστόσο μπορεί να

Σχήμα 2.1: Τα κύρια στοιχεία ενός συστήματος ενισχυτικής μάθησης

χειροτερέψει την επίδοσή της επειδή οι ενέργειες που εκτελεί μπορεί να μην είναι καλύτερες από τις τωρινές. Υπάρχουν και περιπτώσεις όπου το περιβάλλον αλλάζει συνεχώς και ο πράκτορας θα πρέπει να εξερευνεί συνεχώς για να έχει ενημερωμένη την τακτική (policy) του και έτσι θα πρέπει να εξερευνά (exploration) συνεχώς αλλά και να εκμεταλλεύεται τα δεδομένα που έχει (exploitation).

Στην ενισχυτική μάθηση μπορούν να χρησιμοποιηθούν διάφοροι αλγόριθμοι όπως Monte Carlo το οποίο είναι μία ευρεία κλάση από αλγορίθμους υπολογισμών οι οποίοι βασίζονται στην τυχαία δειγματοληψία ώστε να παράγουν αριθμητικά αποτελέσματα. Ένας ακόμα είναι το Q-Learning, είναι ένας αλγόριθμος ενισχυτικής μάθησης όπου μαθαίνει το κόστος μίας ενέργειας σε μία συγκεκριμένη κατάσταση. Ένας ακόμη είναι ο Deep Q-Learning όπου η διαφορά με το Q-Learning είναι ότι δεν χρησιμοποιεί ένα Q-table ώστε να ταιριάξει μια κατάσταση με μία ενέργεια αλλά χρησιμοποιεί ένα δίκτυο νευρώνων το οποίο χαρτογραφεί τις καταστάσεις στην είσοδο του δικτύου με το ζευγάρι ενέργεια και Q-value που προκύπτουν στην έξοδό του, το οποίο εφαρμόζεται στην τεχνική Deep Reinforcement Learning [7].

2.2 Q-Network στην ενισχυτική μάθηση (Deep Reinforcement Learning)

Το Q-learning είναι ένας αλγόριθμος ενισχυτικής μάθησης ο οποίος δεν απαιτεί ένα μοντέλο του περιβάλλοντος ώστε να βρει την καλύτερη ενέργεια που χρειάζεται να λάβει ο πράκτορας σε κάθε περίπτωση. Μπορεί να χειριστεί προβλήματα με στοχαστικές μεταβάσεις και αποτιμήσεις χωρίς να χρειάζεται να απαιτεί προσαρ-

μογές.

Αν έχουμε ένα πεπερασμένο αριθμό διαδικασιών απόφασης Markov ο αλγόριθμος αυτός μπορεί να βρει τη βέλτιστη τακτική στοχεύοντας πάντα στη μεγιστοποίηση της ανταμοιβής σε κάθε επόμενη ενέργεια από αυτή που βρίσκεται κάθε φορά. Ο αλγόριθμος αυτό μπορεί να βρεί τη βέλτιστη τακτική για την επιλογή ενεργειών για κάθε πεπερασμένο αριθμό διαδικασιών απόφασης Markov αν έχει στη διάθεσή του άπειρο χρόνο για εξερεύνηση [8].

Το γράμμα Q στον αλγόριθμο αντιστοιχεί στο όνομα της συνάρτησης που χρησιμοποιεί ο αλγόριθμος για να υπολογίσει την ανταμοιβή(reward) που προκύπτει από μία ενέργεια του πράκτορα σε κάθε δεδομένη κατάσταση του περιβάλλοντος. Η μέγιστη τιμής της συνάρτησης αυτής προκύπτει από την αρχική παρατήρηση s, την ενέργεια a και τη βέλτιστη τακτική (policy). Η εξίσωση της συνάρτησης αυτή είναι:

$$Q*(s,a) = r(s,a) + maxQ(s', a)$$

Η τιμή της εξίσωσης αυτής ονομάζεται Q-Value και είναι το άθροισμα της ανταμοιβής μίας ενέργειας a στην κατάσταση s και της μεγαλύτερης δυνατής τιμής Q από την επόμενη κατάσταση. Το γ είναι ο συντελεστής έκπτωσης ο οποίος ελέγχει την κατανομή των αποτιμήσεων μελλοντικά στο πρόβλημα. Η τιμή του Q(s',a) δηλαδή η τιμή Q-value της επόμενης ενέργειας από την Q(s,a) εξαρτάται από την τιμή του Q(s'',a) οπότε επιλέγοντας την κατάλληλη τιμή του γ μπορούμε να ελέγξουμε κατά πόσο οι επόμενες τιμές της ανταμοιβής (reward-Q(s'',a)) θα επηρεάζουν ή δεν θα επηρεάζουν το αποτέλεσμα της ανταμοιβής (Q(s',a)) στην τωρινή κατάσταση.

Ο αλγόριθμος Q-learning είναι ένας απλός αλλά ισχυρός αλγόριθμος για τη δημιουργία ενός πίνακα που καθοδηγεί τον πράκτορα ώστε να λαμβάνει τις σωστές αποφάσεις. Ωστόσο όμως ένα περιβάλλον μπορεί να αποτελείται από πάνω από 20 χιλιάδες καταστάσεις και 2 χιλιάδες ενέργειες σε κάθε κατάσταση, αυτό θα έχει σαν αποτέλεσμα τη δημιουργία ενός πίνακα που θα περιέχει πάνω από 20 εκατομμύρια κελιά. Καθώς θα αποθηκεύονται νέα δεδομένα και θα ενημερώνονται τα παλιά ο πίνακας θα μεγαλώνει σε μέγεθος. Για αυτό το λόγο προτάθηκε η ιδέα της προσέγγισης των Q-values χρησιμοποιώντας μοντέλα της μηχανικής μάθησης όπως

είναι τα νευρωνικά δίκτυα.

Η τεχνική αυτή ονομάζεται deep Q-learning όπου χρησιμοποιούμε ένα δίκτυο νευρώνων ώστε να προσεγγίζουμε τη συνάρτηση Q-value. Το δίκτυο νευρώνων λαμβάνει ως είσοδο την κατάσταση του περιβάλλοντος που βρισκόμαστε και δίνει ως έξοδο όλες τις τιμές της Q-value από όλες τις δυνατές ενέργειες του πράκτορα. Ο τρόπος με τον οποίο λειτουργεί είναι ότι αρχικά γίνεται αποθήκευση προηγούμενων εμπειριών στη μνήμη του χρήστη, η επόμενη ενέργεια είναι η μεγαλύτερη τιμή που θα προκύψει από το Q-δίκτυο (Q-Network) και η συνάρτηση απωλειών είναι η μέση τιμή του τετραγωνισμένου λάθους που έχει προκύψει από την προσεγγισμένη τιμή Q-value συν της Q-value που στοχεύουμε μείον του Q* (μέγιστη τιμής της συνάρτησης).

$$Q(St, At)□Q(St, At) + a[Rt + 1 + maxQ(St + 1, a)–Q(St, At)]$$

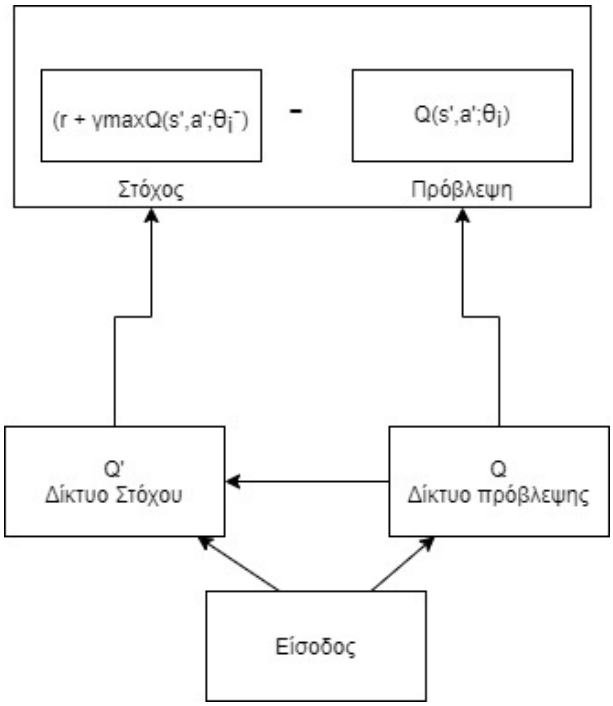
Όπου το R είναι η αληθινή αντικειμενική ανταμοιβή. Το δίκτυο ενημερώνει τις τιμές χρησιμοποιώντας οπισδοδιάδοση (back-propagation) ώστε τελικά να συγκλίνει.

Στην ενισχυτική μάθηση η τακτική ή οι συναρτήσεις τιμών που χρησιμοποιούνται για να λάβουμε τις ενέργειες που θα εκτελέσει ο πράκτορας αλλάζουν συνεχώς καθώς σε κάθε επανάληψη μαθαίνουμε περισσότερες τιμές για τις καταστάσεις και τις ενέργειες και με αυτό τον τρόπο αλλάζει και το αποτέλεσμα. Προσπαθούμε συνεχώς να συνδέσουμε τις συχνά αλλαγμένες εισόδους με τις συχνά αλλαγμένες εξόδους. Αυτό γίνεται με τη βοήθεια των δικτύων στόχου (Target Network) και της επανάληψης εμπειρίας (Experience Replay).

Στο δίκτυο στόχου χρησιμοποιούμε δύο νευρωνικά δίκτυα όπου το ένα χρησιμοποιείται για την εκμάθηση και το δεύτερο για την εκτίμηση του στόχου. Το δίκτυο του στόχου παίρνει σε κάθε επανάληψη τις παραμέτρους του δικτύου πρόβλεψης. Αυτό έχει ως αποτέλεσμα μία πιο σταθερή εκπαίδευση επειδή η συνάρτηση στόχου είναι για λίγο σταθερή. Το experience replay αποθηκεύει σε ένα πίνακα τα δεδομένα τα οποία έχει ανακαλύψει ο αλγόριθμους κατά τη διάρκεια της εκπαίδευσης. Η δομή που έχει το δίκτυο σε εικονική απεικόνιση φαίνεται στο Σχήμα 2.2.

Αν συνδέσουμε όλες τις έννοιες που έχουμε πει μέχρι τώρα μπορούμε να σχημα-

Σχήμα 2.2: Δομή του δικτύου Q-Network



τίσουμε τη διαδικασία με την οποία λειτουργεί ένα deep Q-network. Αρχικά προ-επξεργαζόμαστε τα δεδομένα της αρχικής κατάστασης του προβλήματός μας (state s) στο δίκτυο και έτσι θα λάβουμε ως αποτέλεσμα τις Q-values από όλες τι δυνατές ενέργειες που μπορούν να εκτελεστούν. Έπειτα επιλέγουμε μία τυχαία ενέργεια η οποία θα μας δίνει τη μέγιστη τιμή Q. Στη συνέχεια εκτελούμε την ενέργεια, αποθηκεύουμε το αποτέλεσμα στο replay buffer και υπολογίζουμε τις απώλειες που έχουμε. Αφού εφαρμόσουμε Κυρτή Βελτιστοποίηση (gradient descent) ώστε να μειώσουμε τις απώλειες, αντιγράφουμε τα βάρη του δικτύου πρόβλεψης στο δίκτυο στόχου και επαναλαμβάνουμε τη διαδικασία για αριθμό επαναλήψεων που ορίζει κάθε φορά το πρόβλημα [9].

2.3 Replay Buffer-Experience Replay (επανάληψη εμπειρίας)

Η επανάληψη εμπειρίας χρησιμοποιείται ευρέως σε αλγόριθμους βαθιάς ενισχυτικής μάθησης (deep reinforcement learning). Είναι η μόνη μέθοδος έως τώρα η οποία μπορεί να παράξει δεδομένα τα οποία δεν είναι συνδεδεμένα μεταξύ τους και επιπλέον βελτιώνει σημαντικά την αποδοτικότητα των δεδομένων για τα συστήματα ενισχυτικής μάθησης τα οποία συνήθως έχουν μεγάλη ανάγκη για δεδομένα. Με τη χρήση αυτής της τεχνικής αποθηκεύουμε τις εμπειρίες του πράκτορα σε κάθε

κατάσταση. Η αποθηκευμένη εμπειρία περιλαμβάνει την κατάσταση του περιβάλλοντος, την ενέργεια που εκτέλεσε στην κατάσταση που βρίσκεται και την ανταμοιβή (reward) που έλαβε ο πράκτορας για την ενέργεια που εκτέλεσε στη συγκεκριμένη χρονική στιγμή και το αποτέλεσμα που προκάλεσε στο περιβάλλον του.

Ο κύριος λόγος που χρησιμοποιούμε μνήμη επανάληψης (replay memory) είναι ώστε η εκπαίδευση του δικτύου να μην γίνεται από δείγματα τα οποία σχετίζονται μεταξύ τους. Αν το δίκτυο εκπαιδευόταν από συνεχόμενα δείγματα εμπειρίας καθώς συμβαίνουν με τη σειρά στο περιβάλλον, το δίκτυο θα εκπαιδευόταν με στοιχεία τα οποία έχουν μεγάλη σχέση μεταξύ τους και αυτό θα οδηγούσε σε μία ανεπαρκής εκπαίδευση. Ωστόσο αν παίρνουμε τυχαία δείγματα από τη μνήμη εμπειρίας τότε θα σπάσουμε αυτό τον συσχετισμό και θα έχουμε καλύτερα αποτελέσματα [10].

Το μέγεθος του replay buffer παίζει σημαντικό ρόλο στη διαδικασία της εκπαίδευσης. Η επιλογή του μεγέθους του έχει υποτιμηθεί από την κοινότητα που ασχολείται με ενισχυτική μηχανική μάθηση αλλά έχει αποδειχθεί ότι το μέγεθος είναι μία παράμετρος που χρειάζεται προσεκτική προσέγγιση. Το μέγεθος του replay buffer εξαρτάται από τη συνάρτηση που θα χρησιμοποιήσουμε για να αναπαραστήσουμε το περιβάλλον και για ποιον αλγόριθμο. Για παράδειγμα αν χρησιμοποιήσουμε μία μη γραμμική συνάρτηση ο αλγόριθμος Online-Q λειτουργεί καλύτερα με μικρού μεγέθους replay buffer ενώ ο αλγόριθμος Buffer-Q λειτουργεί καλά με μεσαίου μεγέθους replay buffer, αλλά δεν μπορεί να εξάγει αποτέλεσμα για πολύ μεγάλου μεγέθους replay buffer. Για συνάρτηση πίνακα ένα replay buffer με μεγάλο μέγεθος είναι αρκετό και διατηρεί όλες τις αλλαγές του περιβάλλοντος για τους αλγορίθμους Online-Q και Buffer-Q [11].

2.4 Τακτική (policy) στην ενισχυτική μάθηση

Στη ενισχυτική μάθηση η τακτική είναι ένα από τα κυριότερα μέρη η οποία επιτρέπει στον πράκτορα να λαμβάνει τις σωστές αποφάσεις καθώς αλληλοεπιδρά με το περιβάλλον.

Η τακτική του πράκτορα είναι η σύνδεση όλων των πιθανών ενεργειών με όλες τις πιθανές καταστάσεις. Μπορεί να είναι σε μορφή ενός πίνακα όπου ψάχνει εκεί ο πράκτορας ή ο πράκτορας μπορεί να βασίζεται σε μία τακτική η οποία είναι υπολογισμένη μέσω της αναζήτησης σε μια πληθώρα τιμών οι οποίες μεταβάλλονται

συνεχώς και χρησιμοποιώντας το μοντέλο του περιβάλλοντος αναζητά μία σειρά ενεργειών οι οποίες παράγουν το καλύτερο δυνατό αποτέλεσμα [12].

Αρχικά, ο πράκτορας δεν έχει διαθέσιμη μία καλή τακτική η οποία του δίνει τη μέγιστη τιμή ανταμοιβή ή τον βοηθά να πετύχει τον στόχο του. Η εκμάθηση μία καλής τακτικής από τον πράκτορα βασίζεται σε μία επαναληπτική διαδικασία όπου η τακτική αρχικά αρχικοποιείται τυχαία, δηλαδή ο πράκτορας λαμβάνει μία τυχαία ενέργεια σε κάθε κατάσταση και έπειτα επαληθεύει εάν αυτές οι ενέργειες τον οδήγησαν σε θετική ή αρνητική ανταμοιβή (reward). Έπειτα από πολλές επαναλήψεις ο πράκτορας αρχίζει να μαθαίνει ποια είναι η τακτική η οποία μπορεί να δώσει θετική ανταμοιβή και να τον οδηγήσει στόχο. Αν φτάσει σε αυτό το σημείο σημαίνει ότι έχει βρει τη βέλτιστη τακτική.

Στην ενισχυτική μάθηση υπάρχουν δύο μορφές τακτικής η ντετερμινιστική και η στοχαστική. Στη ντετερμινιστική τακτική υπάρχει μία μόνο συγκεκριμένη ενέργεια για μία δεδομένη κατάσταση. Μόλις ο πράκτορας φτάσει σε μία δοθείσα κατάσταση τότε η ντετερμινιστική τακτική λέει στον πράκτορα πάντα να εκτελέσει μία συγκεκριμένη ενέργεια. Αντιθέτως μία στοχαστική τακτική επιστρέφει μία κατανομή πιθανοτήτων από πολλαπλές ενέργειες που μπορούν εκτελεστούν σε μία κατάσταση. Οπότε ο πράκτορας μπορεί να εκτελεί διαφορετικές ενέργειες κάθε φορά που επισκέπτεται μία συγκεκριμένη κατάσταση, βασισμένες στις τιμές που έχουν οι ενέργειες στην κατανομή πιθανοτήτων τους, που έχει επιστρέψει η τακτική [13].

Στην ενισχυτική μηχανική μάθηση μπορούμε να συναντήσουμε αλγόριθμους που περιγράφονται ως On-policy και άλλους που περιγράφονται ως Of-policy. Οι αλγόριθμοι On-policy λύνουν το δίλημμα μεταξύ εκμετάλλευσης και εξερεύνησης αφού περιλαμβάνει τυχαιότητα στην τακτική. Επιλέγουν τυχαίες ενέργειες αρχικά και έπειτα επιλέγουν τη βέλτιστη ενέργεια. Προσπαθούν συνεχώς να εκτιμήσουν ή να βελτιώσουν την τακτική που χρησιμοποιείται [14]. Επειδή όμως θα επιλέγεται πολύ συχνά η βέλτιστη ενέργεια ο πράκτορας ρισκάρει να καταλήξει σε τοπικό βέλτιστο. Ένα παράδειγμα On-policy αλγορίθμου είναι ο SARSA όπου επιλέγει μία ενέργεια, λαμβάνει μία ανταμοιβή (reward) και έπειτα μεταφέρεται σε μία επόμενη κατάσταση για να εκτελέσει την επόμενη ενέργεια. Από την άλλη πλευρά οι αλγόριθμοι Off-policy προσφέρουν άλλη λύση για το πρόβλημα της εξερεύνησης και εκμετάλλευσης χρησιμοποιώντας δύο τακτικές μία για τη συμπεριφορά και μία για τον

στόχο. Η τακτική συμπεριφοράς χρησιμοποιείται για την εξερεύνηση και για δημιουργία επεισοδίων και η τακτική στόχου χρησιμοποιείται για την εκτίμηση και βελτίωση των συναρτήσεων. Η μέθοδος αυτή είναι αποτελεσματική, ωστόσο υπάρχει μία αναντιστοιχία μεταξύ αυτών που προσπαθούμε να εκτιμήσουμε και αυτών που λαμβάνουμε δείγματα. Για αυτό το λόγο χρησιμοποιείται μία τεχνική που ονομάζεται δειγματοληψία σπουδαιότητας (importance sampling) ώστε να λυθεί αυτή η αναντιστοιχία. Ένας γνωστός αλγόριθμος off-policy είναι ο Q-learning όπου κάθε φορά προσπαθούμε να βελτιώσουμε τη συνάρτηση επιλέγοντας συνεχώς την ενέργεια που επιλέγει τη μέγιστη τιμή [15].

2.5 Εκμετάλλευση και εξερεύνηση (exploration vs exploitation)

Στην ενισχυτική μάθηση ο πράκτορας συνεχώς εξετάζει το περιβάλλον πολλές φορές ώστε να βρει τις βέλτιστες ενέργειες. Η εξισορρόπηση μεταξύ εξερεύνησης και εκμετάλλευσης είναι αρκετά σημαντική επειδή ο πράκτορας μπορεί να έχει βρει ένα καλό μονοπάτι αλλά μπορεί να υπάρχει ένα καλύτερο μονοπάτι το οποίο δεν έχει βρει ακόμα. Αν ο πράκτορας δεν εξερευνά τότε θα βρίσκει το πρώτο καλό μονοπάτι που συναντά και ένα καλύτερο που υπάρχει δεν θα βρεθεί ή ο στόχος μπορεί να βρίσκεται σε μία περιοχή όπου η ανταμοιβή (reward) είναι πολύ μικρή και χωρίς εξερεύνηση δεν πρόκειται να βρεθεί. Αν όμως ο πράκτορας κάνει μεγάλη εξερεύνηση δεν θα καταφέρει να ακολουθήσει ένα μονοπάτι, δεν θα γίνει εκμάθησή του και δεν θα εκμεταλλευτεί τις γνώσεις που αποκτά. Για αυτό το λόγο είναι σημαντικό να εξασφαλιστεί ισορροπία μεταξύ της εξερεύνησης και της εκμετάλλευσης.

Η εξερεύνηση στην ενισχυτική μάθηση είναι η εκτέλεση ενεργειών από τον πράκτορας χωρίς να ξέρει ποιες είναι οι κατάλληλες και τις δοκιμάζει προσπαθώντας να βρει ποιες έχουν ως αποτέλεσμα την καλύτερη ανταμοιβή (reward). Αντίθετα η εκμετάλλευση προσπαθεί να χαρτογραφήσει τις σχέσεις μεταξύ των καταστάσεων και των ενεργειών ώστε να μεγιστοποιηθεί η ανταμοιβή σε ένα άγνωστο και αβέβαιο περιβάλλον [16].

Το δίλημμα μεταξύ επιλογής εξερεύνησης και εκμετάλλευσης για έναν αλγόριθμο γίνεται αρκετά απαιτητικό όταν έχουμε να κάνουμε με τις πιθανές καταστάσεις και ενέργειες επειδή όσο αυτές αυξάνονται, αυξάνεται και ο χρόνος που χρειάζεται ο πράκτορας ώστε να αποκτήσει γνώση για το περιβάλλον. Με τη χρήση κάποιων

αλγορίθμων όπως ο E3 αλγόριθμος, R-Max και REX προσπαθούμε να μετριάσουμε κατά πόσο θα εξερευνούμε και πότε θα αρχίζουμε να εκμεταλλευόμαστε τα δεδομένα χωρίς όμως να υπάρχουν ενέργειες για τις οποίες δεν έχει γίνει εξερεύνηση για την επίδοσή τους [17].

2.6 Βελτιστοποίηση και ο αλγόριθμος Adam

Η μηχανική μάθηση περιλαμβάνει τη χρήση ενός αλγορίθμου με σκοπό την εκμάθηση και γενικοποίηση παλιών δεδομένων από τον πράκτορα ώστε να κάνει νέες προβλέψεις για καινούργια δεδομένα. Αυτή η διαδικασία μπορεί να περιγραφτεί ως μία συνάρτηση εκτίμησης όπου χαρτογραφεί τις εισόδους με τις εξόδους. Για να γίνει αυτή η χαρτογράφηση πρέπει η συνάρτηση εκτίμησης να γίνει μία συνάρτηση βελτιστοποίησης ώστε η μηχανική μάθηση χρησιμοποιώντας έναν αλγόριθμο που παραμετροποιεί τα δεδομένα και έναν αλγόριθμο βελτιστοποίησης να ελαττώσει την τιμή λάθους (error) της συνάρτησης καθώς χαρτογραφεί τις εισόδους με τις εξόδους. Η μείωση αυτής της τιμής του λάθους είναι και ο κύριος στόχος της βελτιστοποίησης. Η βελτιστοποίηση αυτή μπορεί να γίνει με τη χρήση διαφορετικών αλγορίθμων οι οποίοι όμως κάνουν διαφορετικές προβλέψεις κάθε φορά στην αντιστοίχιση των εισόδων και εξόδων [18].

Ένα παράδειγμα αλγόριθμου βελτιστοποίησης είναι ο αλγόριθμος Adam. Είναι μία αποδοτική μέθοδος για στοχαστική βελτιστοποίηση ο οποίος χρησιμοποιεί εκτιμήσεις από την πρώτη και δεύτερη βαθμίδα του νευρωνικού δικτύου ώστε να προσαρμόσει το ποσοστό εκμάθησης για κάθε βάρος του δικτύου. Έτσι κάνει χρήση ελάχιστης μνήμης του συστήματος, κάνοντάς τον πιο αποδοτικό για στοχαστική βελτιστοποίηση. Είναι μία επέκταση του αλγορίθμου gradient descent (καθοδικής κλίσης). Είναι ένας συνδυασμός δύο μεθόδων καθοδικής κλίσης των Momentum και RMSP [19]. Είναι μία μέθοδος η οποία συγκλίνει γρήγορα, διορθώνει τον ρυθμό εκμάθησης που εξαφανίζεται αλλά έχει μεγάλο υπολογιστικό κόστος [20].

2.7 Ένας πράκτορας και πολλαπλοί πράκτορες (single-agent and multi-agents)

Ένα δυσκολότερο πρόβλημα από αυτό της εκμάθησης ενός πράκτορα σε ένα περιβάλλον είναι η εκμάθηση πολλαπλών πρακτόρων στο τι ενέργειες εκτελούν καθώς αλληλοεπιδρούν μεταξύ τους. Η χρήση πολλαπλών πρακτόρων αποδεικνύεται συνεχώς ότι είναι αναγκαία για πολλές εφαρμογές. Το πρόβλημα ωστόσο εμφανίζεται κατά την εκμάθηση όπου όταν στο περιβάλλον δραστηριοποιόταν μόνο ένας πράκτορας το περιβάλλον θεωρούταν σταθερό, τώρα όμως με τους πολλαπλούς πράκτορες αυτό δεν μπορεί να συμβεί επειδή οι ίδιοι οι πράκτορες μεταβάλλουν το περιβάλλον [21].

Ένας μονός πράκτορας αλληλοεπιδρά με το περιβάλλον του, εκτελεί κάποια ενέργεια και έπειτα συλλέγει τα αποτελέσματα αυτής της ενέργειας όπως σε ποια νέα κατάσταση βρέθηκε ή αν πέτυχε κάποιο στόχο. Στόχος του πάντα είναι να βελτιστοποιήσει την ανταμοιβή από τις ενέργειές του. Οι πολλαπλοί πράκτορες δουλεύουν με ακριβώς τον ίδιο τρόπο απλώς υπάρχουν πολλοί πράκτορες που αποφασίζουν για τις ενέργειες. Το διαφορετικό στους πολλαπλούς πράκτορες είναι ότι η ενέργεια του καθενός έχει διαφορετικό αποτέλεσμα το οποίο εξαρτάται και από τι ενέργειες που εκτελούν οι άλλοι πράκτορες. Για να λυθούν προβλήματα με πολλούς πράκτορες έχουμε εστιάσει στη θεωρία των παιχνιδιών όπου οι λύσεις προβλημάτων περιλαμβάνουν συμβιβασμούς και συνεργασία.

Οι αλγόριθμοι οι οποίοι χρησιμοποιούν πολλαπλούς πράκτορες χωρίζονται σε δύο κατηγορίες global-level μέθοδος συντονισμού και neighborhood-level μεθόδους συντονισμού. Στις global-level μεθόδους όλοι οι πράκτορες μοιράζονται τις ίδιες ενέργειες και παρατηρήσεις (observations) και υπάρχει ένας εικονικός πράκτορας ο οποίος εκπαιδεύει μια κεντρική συνάρτηση τιμής για όλους του πράκτορες. Από την άλλη πλευρά στη neighborhood-level μέθοδο η συνεργασία μεταξύ των πρακτόρων γίνεται μόνο στις γειτονιές του δηλαδή οι πράκτορες μοιράζονται τις παρατηρήσεις και ενέργειες μόνο με πράκτορες γείτονές τους. Σε κάθε γειτονιά πρακτόρων υπάρχει και ένας εικονικός πράκτορας. Και οι δύο αυτές μέθοδοι έχουν ως σκοπό να λύσουν το πρόβλημα της συνεργασίας σε γενικό επίπεδο και όχι για τον κάθε πράκτορα ξεχωριστά. Σε πολλά πραγματικά σενάρια οι διάφοροι πράκτορες παίζουν

διαφορετικό ρόλο στο περιβάλλον και δεν μπορούν έτσι απλά να συμπεριληφθούν σε ένα εικονικό πράκτορα. Οπότε είναι σημαντικό να γίνει ανάλυση των σχέσεων μεταξύ των πρακτόρων ώστε να δημιουργηθεί μία στρατηγική συντονισμού των πρακτόρων [22].

2.8 Monte Carlo

Η μέθοδος του Monte Carlo είναι αρκετό καιρό γνωστή ως ένας δραστικός τρόπος για την εκτέλεση συγκεκριμένων υπολογισμών οι οποίοι κανονικά είναι αρκετά πολύπλοκοι για μία κλασική προσέγγιση. Μία τεχνική λέγεται Monte Carlo όταν χρησιμοποιεί τυχαίους αριθμούς για να λύσει ένα πρόβλημα. Βασίζεται σε συχνές δειγματοληψίες και στατιστικές αναλύσεις ώστε να υπολογίσει τα αποτελέσματα. Δεν έχουμε γνώση του περιβάλλοντος παρά μόνο τις ανταμοιβές (reward) όπου ο Monte Carlo μαθαίνει κάθε φορά παίρνοντας το μέσο όρο από αυτές τις ανταμοιβές [23]. Για κάθε είσοδο στην προσομοίωση του Monte Carlo ορίζουμε μία στατιστική κατανομή η οποία είναι και η πηγή των παραμέτρων εισόδου. Έπειτα παίρνουμε τυχαία δείγματα από κάθε κατανομή τα οποία είναι και οι τιμές των παραμέτρων εισόδου. Για κάθε ένα πακέτο από παραμέτρους εισόδου έχουμε και ένα πακέτο παραμέτρων εξόδου. Κάθε έξοδος έχει και διαφορετική τιμή η οποία αντιστοιχεί και σε ένα συγκεκριμένο αποτέλεσμα στην προσομοίωση. Γίνεται συλλογή των τιμών αυτών, βγαλμένες από αρκετές προσομοιώσεις και τέλος εκτελούνται στατιστικές αναλύσεις των τιμών ώστε να παρθούν αποφάσεις για τις μετέπειτα ενέργειες του πράκτορα.

Αρχικά σε μία Monte Carlo προσομοίωση δημιουργείται ένα ντετερμινιστικό σενάριο το οποίο αντικατοπτρίζει και το πραγματικό σενάριο και όπου χρησιμοποιούμε και τις πιθανές τιμές των παραμέτρων εισόδου. Όταν έχουμε δημιουργήσει ένα ικανοποιητικό ντετερμινιστικό μοντέλο προσθέτουμε τις συνιστώσες κινδύνου οι οποίες προκύπτουν από τη στοχαστική φύση των μεταβλητών εισόδου. Έτσι γίνεται αναγνώριση της κατανομής εισόδου. Έπειτα για τις μεταβλητές εισόδου παράγουμε μία σειρά από τυχαίους αριθμούς για τις κατανομές που δημιουργήσαμε προηγουμένως. Κάθε σειρά από τυχαίους αριθμούς περιλαμβάνει μία τιμή για κάθε μία από τις παραμέτρους εισόδου οι οποίες χρησιμοποιούνται στο ντετερμινιστικό μοντέλο ώστε να δημιουργεί μία σειρά από τιμές εξόδου. Η διαδικασία αυτή επαναλαμβά-

νεται για όλες τις κατανομές εισόδου. Τέλος γίνεται η ανάλυση και η λήψη αποφάσεων. Αφού έχουν συλλεχθεί οι τιμές των εξόδων από την προσομοίωση εκτελούμε στατιστική ανάλυση των τιμών αυτών. Με την εκτέλεση αυτού του βήματος έχουμε στατιστική εικόνα για τις ενέργειες που θα εκτελέσει ο πράκτορας έπειτα από τη προσομοίωση του Monte Carlo [24].

2.9 Ρυθμός εκμάθησης (learning rate)

Τα βάρη που περιέχονται σε ένα δίκτυο νευρώνων δεν μπορούν να υπολογιστούν χρησιμοποιώντας κάποια αναλυτική μέθοδο, αλλά τα βάρη πρέπει να καθοριστούν μέσω μίας εμπειρικής βελτιστοποίησης. Η βελτιστοποίηση αυτή είναι απαιτητική και το βάρος που έχει ο κάθε κόμβος του νευρωνικού δικτύου μπορεί να επηρεαστεί από λύσεις που φαίνονται καλές όπως είναι τα τοπικά μέγιστα ή να καταλήξει σε λύσεις που είναι εύκολο να βρεθούν όπως είναι τα τοπικά ελάχιστα. Ένα δίκτυο νευρώνων μαθαίνει ή προσεγγίζει μία συνάρτηση ώστε να αντιστοιχίσει τις εισόδους με τις εξόδους στο σύνολο των δεδομένων της εκπαίδευσης. Το ποσοστό μάθησης (learning rate) είναι υπεύθυνο για τον έλεγχο του ρυθμού ή της ταχύτητας με την οποία το δίκτυο μαθαίνει. Μεταβάλλει τα βάρη του δικτύου ελέγχοντας το λάθος (error) που έχουν τα βάρη αυτά στο μοντέλο. Εάν έχουμε ένα τέλειο ρυθμό εκμάθησης το μοντέλο θα μάθει με μεγάλη ακρίβεια να αντιστοιχεί εισόδους με εξόδους. Πρέπει ωστόσο να γίνεται έλεγχος για το μέγεθος του ρυθμού εκμάθησης επειδή ένας μεγάλος ρυθμός εκμάθησης επιτρέπει στο μοντέλο να μαθαίνει πιο γρήγορα αλλά θα κάνει μεγάλες ενημερώσεις στα βάρη και η επίδοση του μοντέλου θα ταλαντεύεται μεταξύ των διάφορων μορφών του περιβάλλοντος. Αν είναι μικρό σε μέγεθος τότε επιτρέπουμε στο μοντέλο να έχει βάρη με εκμάθηση πιο βέλτιστη από ότι θα είχε από ένα μεγάλο ρυθμό εκμάθησης αλλά η εκμάθηση αυτή θα διαρκέσει αρκετά περισσότερο [25].

2.10 Ρυθμός έκπτωσης (discount rate)

Σε ένα μοντέλο ενισχυτικής μηχανικής μάθησης ο ρυθμός έκπτωσης αποφασίζει για τα αποτελέσματα της εκμάθησης. Αν δεν υπάρχει ένας καλός ρυθμός έκπτωσης τότε ο πράκτορας θα εστιάζει μόνο στις κοντινές μικρές ανταμοιβές και όχι

στις μελλοντικές μεγάλες ανταμοιβές που θα πετύχει στο μέλλον. Με τη χρήση του ρυθμού έκπτωσης προσπαθούμε να οριοθετήσουμε τις ανταμοιβές σε κάθε βήμα. Αλλάζοντας την τιμή του μπορούμε να κάνουμε το πράκτορα να εστιάσει σε πιο μακρινούς στόχους. Για παράδειγμα αν τιμή γ του ρυθμού έκπτωσης γίνει μηδέν τότε ο στόχος του πράκτορα γίνεται η μεγιστοποίηση της άμεσης ανταμοιβής που έχει λάβει. Καθώς η τιμή του γ αυξάνεται ο πράκτορας αρχίζει να γίνεται πιο προνοητικός και δίνει μεγαλύτερη σημασία στις επερχόμενες ανταμοιβές. Αν έχουμε όμως μία μεγάλη τιμή του γ αυτό μας οδηγεί στο να έχουμε μία αργή σύγκλιση της συνάρτησης τιμών, το οποίο οδηγεί και σε αργό ρυθμό εκμάθησης. Οπότε για να έχουμε μία αποδοτική εκμάθηση θα πρέπει να επιλέξουμε μία μικρή τιμή του γ σε λογικά πλαίσια πάντα και να μην επιλέξουμε στα τυφλά απλώς μία μεγάλη τιμή του γ [26].

Κεφάλαιο 3

Εφαρμογές Ενισχυτικής Μάθησης στον Μηχανοκίνητο Αθλητισμό

3.1 Βελτιστοποίηση της αεροδυναμικής γεωμετρίας στην F1 με μηχανική μάθηση

Η οργάνωση της Formula 1 δεν άργησε να καταλάβει τις δυναμικές των συστημάτων AWS και όπως παρουσιάζουν ερευνητές της AWS[27] στο blog τους, η F1 με τη βοήθειά τους βελτιστοποιούν τις αεροδυναμικές γεωμετρίες μέσω σχεδιασμών πειραμάτων και μηχανικής μάθησης.

Τα αυτοκίνητα της F1 είναι τα πιο γρήγορα στον κόσμο και μπορούν διέρχονται από στροφές με μεγάλες ταχύτητες λόγω των υψηλών επιπέδων αεροδυναμικής που παράγουν. Κάθε ομάδα της F1 είναι υπεύθυνη στο να παράξει ένα όχημα το οποίο θα ταιριάζει με του κανονισμούς του αθλήματος. Αντί όμως οι ομάδες να ξοδεύουν μεγάλα ποσά για να δοκιμές στην πίστα και δοκιμές σε αεροδυναμικές σήραγγες χρησιμοποιούν Υπολογιστική Ρευστοδυναμική όπου τους παρέχει τη δυνατότητα να δημιουργούν ένα εικονικό περιβάλλον όπου μελετούν τη ροή του αέρα γύρω από τα μέρη του οχήματος. Για να μπορέσουν να κάνουν ακόμα καλύτερα τα αποτελέσματα των δοκιμών τους η F1 συνεργάστηκε με την AWS ώστε με τη βοήθεια της μηχανικής μάθησης να δημιουργήσουν μία ροή πειραμάτων με μηχανική μάθηση που θα συμβουλεύει τους αεροδυναμιστές της F1 ποια σχέδια να εξετάζουν για να μεγιστοποιήσουν τις επιδόσεις των κομματιών του αυτοκινήτου και τις γνώσεις τους.

Οι κλασικές μέθοδοι που ακολουθούν οι αεροδυναμιστές περιλαμβάνουν ένα

χώρο γεμάτο δεδομένα από όλα τα σχέδιά τους ώστε να μπορέσουν στο τέλος να τα συσχετίσουν μεταξύ τους με αποτέλεσμα να χρειάζεται να κάνουν πολλές προσομοιώσεις για να γίνει αυτή η συσχέτιση. Τα επαναληπτικά μοντέλα της μηχανικής μάθησης χρησιμοποιούν τα αποτελέσματα από προηγούμενες προσομοιώσεις για να προβλέψουν την αεροδυναμική αντίδραση τους, καθώς και να δώσουν μία ένδειξη για τη σημαντικότητα της κάθε παραμέτρου του σχεδιασμού.

Η επιλογή για το ποια παράμετρος θα εξετασθεί επόμενη απαιτεί προσεκτική αντιμετώπιση επειδή ο αεροδυναμιστής μπορεί να εκμεταλλευτεί παραμέτρους που του έχει προβλέψει το μοντέλο μηχανικής μάθησης και τα οποία παρέχουν υψηλή αεροδυναμική με κόστος όμως την αποτυχή εξερεύνηση περιοχών του χώρου σχεδίασης οι οποίες παρέχουν ακόμη μεγαλύτερη αερροδυναική. Για αυτό το λόγο εξετάζονται τρεις μορφές αποτελεσμάτων μία που βασίζεται στη δυναμική και δύο που βασίζονται στην εξερεύνηση.

Δημιουργήθηκε ένα μοντέλο το οποίο μετράει τη σχετική σημασία της κάθε παραμέτρου υπολογίζοντας την αύξηση του σφάλματος αφού ανακατέψει τις τιμές της κάθε παραμέτρου. Αν μία παράμετρος είναι σημαντική αυξάνεται το σφάλμα πρόβλεψης και έτσι ο αεροδυναμιστής δίνει μεγαλύτερο βάρος σε αυτή .

3.2 Μηχανική μάθηση για τη μοντελοποίηση και ανάλυση αγώνων οδήγησης

Ο Alexander [28] παρουσίασε τη μοντελοποίηση των ανθρώπινων χαρακτηριστικών οδήγησης και σε συνδυασμό με τα χαρακτηριστικά των οχημάτων που οδηγούν να στοχεύσουν στην καλύτερη δυνατή θέση στον αγώνα.

Ομάδες στον μηχανοκίνητο αθλητισμό συνεχώς προσπαθούν να κατασκευάσουν το πιο γρήγορο αυτοκίνητο που ξεπερνάει όλους τους ανταγωνιστές τους. Είναι μία δύσκολη αποστολή που στοχεύει στη μείωση του μέσου όρου των γύρων λαμβάνοντας υπόψιν αρκετές παραμέτρους. Οι μηχανικοί έχουν κατανοήσει και μοντελοποιήσει τις δυναμικές των οχημάτων, ωστόσο η ταχύτητα του οχήματος εξαρτάται και από αυτόν που το οδηγεί ο οποίος παίζει σημαντικό ρόλο στην ολική μορφή του οχήματος. Το όχημα πρέπει προσαρμόζεται στις ανάγκες του οδηγού. Για αυτό το λόγο οι ομάδες χρησιμοποιούν Human-Driver-in-the-Loop προσομοιώσεις όπου

δοκιμάζουν τροποποιημένα εξαρτήματα και συστήματα. Οι προσομοιώσεις τους παρέχουν σημαντικές πληροφορίες για το πως όλες οι αλλαγές που εκτελούν πάνω σε αυτά αντιδρούν με τον οδηγό. Αυτή η τεχνική λειτουργεί μόνο για την ανάπτυξη του οχήματος και δεν είναι χρήσιμη για συνεχόμενη βελτίωση τον χρόνο του γύρου λόγο της μεγάλης χρονικής διάρκειας για την εκτίμησή του και τη μεγάλη χρήση πόρων που απαιτείται. Οπότε απαιτείται ένα επαναστατικό και κοντά στην ανθρώπινη συμπεριφορά μοντέλο το οποίο θα μπορεί να μιμηθεί προσωπικά στιλ οδήγησης κάθε οδηγού σε ένα ολικά εικονικό όχημα προσομοίωσης ώστε να αυξηθεί αποδοτικά ο αριθμός των δοκιμών. Την ίδια στιγμή η μοντελοποίηση της συμπεριφοράς του ανθρώπου που οδηγεί έχει προκλήσεις καθώς το όχημα του επιταχύνει συνεχώς προσεγγίζοντας τα όρια της ισορροπίας και το οποίο άμεσα απαιτεί μία επαναστατική τακτική (policy).

Για τη μίμηση της οδηγικής συμπεριφοράς του ανθρώπου χρησιμοποιήθηκε η μέθοδος Probabilistic Modeling of Driver Behavior (ProMoD). Μετά την επιβλεπόμενη εκπαίδευση του νευρωνικού δικτύου χρησιμοποιούνται ανθρώπινες αναπαραστάσεις που παράγει η τεχνική ProMoD. Γίνεται αξιολόγηση της τακτικής που προκύπτει σε ένα απλό περιβάλλον αγωνιστικής οδήγησης στο OpenaAI Gym. Αποδείχθηκε ο η τεχνική ProMoD παράγει ενέργειες μου μοιάζουν με αυτές των ανθρώπων και είναι πιο σταθερές από αυτές που παράγει η άμεση επιβλεπόμενη εκπαίδευση.

Έγινε ανάπτυξη μίας δομημένης μεθόδου για την ανάλυση ξεχωριστών στιλ οδήγησης και έγινε επέκταση της τεχνικής σε ένα επαγγελματικό προσομοιωτή όπου χρησιμοποιούνται δεδομένα από επαγγελματίες οδηγούς αγώνων. Οι αλγόριθμοι ταυτοποίησης οδηγού (Driver Identification) και μετρικός αλγόριθμος κατάταξης (Metric Ranking Algorithm DIRMA) υπολογίζουν ένα σετ από γύρους και μετρήσεις που χαρακτηρίζουν το στιλ οδήγησης του οδηγού αλλά εξαρτώνται κυρίως στις ενέργειες του οδηγού. Για να διατηρήσει την ερμηνευσιμότητα και να μειώσει την πολυπλοκότητα, η DIMRA μειώνει τον αριθμό των μετρήσεων χρησιμοποιώντας διαδοχική προς τα πίσω επιλογή (SBS) σε συνδυασμό με k-medoids ομαδοποίηση.

Εισάγοντας μία προσέγγιση για τον υπολογισμό των δεδομένων για άγνωστες πίστες μέσω της εμπειρίας από άλλες πίστες, το μοντέλο του οδηγού καταφέρνει να γενικοποιήσει τα δεδομένα από τις πίστες και έχει τη δυνατότητα να δημιουργήσει ανταγωνιστικούς γύρους σε καινούργιες πίστες. Επιπλέον παρουσιάστηκε και μία

διαδικασία προσαρμογής του μοντέλου ώστε να προσαρμόζει τη συμπεριφορά του με βάση τις εμπειρίες από προηγούμενους γύρους. Το μοντέλο είναι ικανό να μάθει από τα λάθη των προηγούμενων γύρων και να πραγματοποιήσει γύρους οι οποίοι δεν έχουν πραγματοποιηθεί προηγουμένως με αυξημένη επίδοση. Με αυτό το μοντέλο έγινε ένα βήμα πιο κοντά στο να γίνει καλύτερη κατανόηση της συμπεριφοράς των ανθρώπινων οδηγών γενικώς και πάντα έχοντας συνείδηση τα προσωπικά χαρακτηριστικά του καθενός. Ακόμα τα αποτελέσματα που προέκυψαν βοηθούν στην ανάπτυξη των αυτόνομων αγώνων αυτοκινήτων. Εκτός από τον αντίκτυπο στους αγώνες οδήγησης το μοντέλο αυτό μπορεί να είναι χρήσιμο για την ανάπτυξη βοηθημάτων οδήγησης σε αυτοκίνητα δρόμου καθώς τα συστήματα αυτά θα μπορούν έχουν ανθρώπινη συμπεριφορά όταν επιδρούν στη συμπεριφορά του οχήματος.

3.3 Από άκρη σε άκρη αγωνιστική οδήγηση με ενισχυτική μάθηση

Οι Jaritz et al. [29] παρουσίασαν στο παγκόσμιο συνέδριο ρομποτικής και αυτοματοποίησης την έρευνά τους για την από άκρη σε άκρη οδήγηση με βαθιά ενισχυτική μάθηση. Με τη χρήση των τελευταίων αλγορίθμων ενισχυτικής μάθησης δημιουργήθηκε μία καινούργια στρατηγική ανταμοιβής και εκμάθησης οι οποίες οδηγούν σε πιο γήγορη προσέγγιση και πιο επαναστατική οδήγηση χρησιμοποιώντας μόνο RGB εικόνες από μια κάμερα. Χρησιμοποιείται μια δομή ασύγχρονης κριτικής ηθοποιού (A3C) ώστε να μάθει τον έλεγχο του αυτοκινήτου σε ένα φυσικό και γραφικά ρεαλιστικό παιχνίδι rally με τους πράκτορες να εξελίσσονται ταυτόχρονα σε πίστες με ποικιλία από δομές δρόμων, σκηνικών και ανωμαλίες του εδάφους. Έγινε απόδειξη μίας πλήρης εκτίμησης και γενικοποίησης σε διαδρομές που δεν έχει συναντήσει ξανά αλγόριθμος, χρησιμοποιώντας νόμιμα όρια ταχύτητας. Η μέθοδος αυτή έδειξε ότι μπορεί να προσαρμοστεί και σε μία ακολουθία από εικόνες της πραγματικότητας.

Προτάθηκε μία μέθοδος η οποία επωφελείται από τις πρόσφατες ασύγχρονες εκμαθήσεις και χτίζεται από την προηγούμενη εργασία εκμάθησης ενός πράκτορα από άκρη σε άκρη (end-to-end) σε ένα στοχαστικό και ρεαλιστικό παιχνίδι αγώνων αυτοκινήτων. Επιπλέον για να παραμείνει κοντά στις πραγματικές συνθήκες οδήγησης δόθηκε βάση μόνο σε εικόνες και ταχύτητα για να προβλεφθεί ο πλήρης έλεγχος κατά μήκος και πλευρικά του οχήματος. Μαζί με τη στρατηγική εκμάθησης

η μέθοδος κάνει πιο γρήγορες εκτιμήσεις από προηγούμενες παρόμοιες ερφαρμογές και προβάλει μία γενικοποίηση. Παρόλο των σημαντικά πιο πολύπλοκου περιβάλλοντος που περιλαμβάνει 29,6 χιλιόμετρα από πίστες για την εκμάθηση της μεθόδου με διάφορες οπτικές εμφανίσεις όπως χιόνια, βουνά και ανωμαλίες του εδάφους, ο αλγόριθμος δοκιμάστηκε με επιτυχία και σε πραγματικό περιβάλλον παρόλο που ήταν εκπαιδευμένο μόνο σε περιβάλλον προσομοίωσης.

3.4 Μηχανική μάθηση για την κατηγοριοποίηση και πρόβλεψη των αντικειμενικών και υποθετικών αξιλογήσεων των δυναμικών κινήσεων ενός οχήματος

Οι Gómez et al. [30] δημοσίευσαν ένα άρθρο που αφορά τη μηχανική μάθηση για την κατηγοριοποίηση και πρόβλεψη των αντικειμενικών και υποθετικών αξιολογήσεων των δυναμικών κινήσεων ενός οχήματος.

Υποκειμενικές μετρήσεις και μηχανικές προσομοιώσεις με τη βοήθεια υπολογιστή δεν μπορούν εκμεταλλευτούν στο μέγιστο καθώς είναι πολύ σημαντική η επίδραση του οδηγού στην εξέλιξη ενός οχήματος. Ακόμη παρόλες τις ενέργειες που έχουν γίνει δεν είναι εύκολο να αναγνωρίσουμε τις σχέσεις μεταξύ των υποκειμενικών και αντικειμενικών μετρικών καθώς οι υποκειμενικές αλλάζουν μεταξύ οδηγών, μεταξύ γεωγραφικών θέσεων και με το χρόνο. Παρουσιάζεται μία έρευνα όπου χρησιμοποιούνται δύο τεχνητά δίκτυα νευρώνων τα οποία είναι χτισμένα το ένα πάνω στο άλλο. Το ένα δίκτυο βασίζεται την αντικειμενική μετρική, παράγει ένα χάρτη ο οποίος ομαδοποιεί μαζί παρόμοια οχήματα και δίνει τη δυνατότητα να οπτικοποιηθεί η ομαδοποίηση των οχημάτων που έχουν μετρηθεί. Αυτός ο χάρτης παρουσιάζει αντικειμενικά ότι υπάρχουν οι οντότητες της μάρκας και του οχήματος. Επίσης προβλέπει τα υποκειμενικά χαρακτηριστικά ενός νέου οχήματος βασισμένο στις απαιτήσεις, στις προσομοιώσεις και στις μετρήσεις που έχουν διεξαχθεί. Αυτά τα χαρακτηριστικά περιγράφονται μέσω της γειτονιάς ενός νέου αυτοκινήτου μέσα στον χάρτη τα οποία δημιουργούνται με βάση γνωστά οχήματα. Αυτή η πρόγνωση επεκτείνεται ώστε να εκτελέσει μία ανάλυση ευαισθησίας των απαιτήσεων καθώς και να επικυρώσει προηγούμενες προτιμώμενες δημοσιευμένες μετρήσεις αίσθησης του τιμονιού. Τέλος οι ποιοτικές πληροφορίες που δίνονται από τις μετρήσεις των

κατηγοριοποιήσεων συμπληρώνονται από ένα δεύτερο δίκτυο νευρώνων. Αυτό το δίκτυο περιγράφει μια επαναληπτική επιφάνεια η οποία δίνει τη δυνατότητα για ποσοτικές προβλέψεις για παράδειγμα την πρόβλεψη της αντικειμενική αίσθησης του τιμονιού ενός καινούργιου οχήματος μέσα από την υποκειμενική του πρόβλεψη.

Η έρευνα αυτή χρησιμοποιεί τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks ANN) ώστε να παράξει μία αυτόματη κατηγοριοποίηση των οχημάτων βασισμένη σε αντικειμενικές μετρικές.Έχουμε μία μη επιβλεπόμενη κατηγοριοποίηση των οχημάτων σε ένα δισδιάστατο χάρτη, ο οποίος παρουσιάζει τα οχήματα σε μία εύκολη στην κατανόηση και πληροφορημένη μορφή δίνοντάς μας τη δυνατότητα για καλύτερη αναγνώριση των ομοιοτήτων που έχουν τα οχήματα μεταξύ τους.

Επιπλέον το δεύτερο τεχνητό νευρωνικό δίκτυο προστίθεται πάνω από τον χάρτη και έχει τη δυνατότητα να προβλέπει υποκειμενικές μετρικές για καινούργια οχήματα με βάση πολλαπλές αντικειμενικές μετρικές, βάζοντάς τους ένα βήμα πιο μπροστά από προηγούμενες μη αποδοτικές έρευνες. Η νέα αυτή μέθοδος επιτρέπει η πρόβλεψη να γίνεται πάνω από τον πίνακα ταξινόμησης ο οποίος είναι μία μέθοδος εξόδου και όχι μόνο μια απλή αξιολόγηση. Επίσης επιτρέπει στους μηχανικούς των οχημάτων να κατανοήσουν πως οι υποκειμενικές μετρικές εξελίσσονται, πως θα τις συνδέσουν με προηγούμενες γνωστές υποκειμενικές μετρήσεις από γνωστά οχήματα κάνοντας χρήση των λέξεων σύννεφα (word-clouds) που περιγράφουν κάθε όχημα στη βάση δεδομένων.

3.5 Χρησιμοποιώντας τη μηχανική μάθηση για να προβλέψουμε αν ένας οδηγός της F1 θα πετύχει πόντους

Ο Pedroso [31] ανέπτυξε ένα πρόγραμμα στο οποίο με τη βοήθεια της μηχανικής μάθησης θέλησε να κάνει προβλέψεις για το αν ένας οδηγός θα πετύχει πόντους σε έναν αγώνα F1.

Στις μέρες μας πολλοί πιστεύουν ότι τα αποτελέσματα στη Formula 1 αποφασίζονται κυρίως από τα χαρακτηριστικά του αυτοκινήτου και όχι από τον οδηγό, το οποίο είναι μερικώς αληθές, αλλά υπάρχουν και άλλοι εξωτερικοί παράγοντες που επηρεάζουν το αποτέλεσμα το οποίοι δίνει κίνητρο στις ομάδες να επικεντρώνονται όλο και περισσότερο στη μηχανική μάθηση.

Αρχικά το πρώτο βήμα ήταν η συλλογή όλων των δεδομένων ώστε να δημιουργηθεί μία νέα συλλογή δεδομένων με αυτά που χρειάζονται. Αφού έγινε αυτή η δημιουργία, έγινε έλεγχος για το πως κατανέμονται οι θέσεις των οδηγών. Έπειτα αφού βρήκαμε ποιοι οδηγοί δεν κατάφεραν να τερματίσουν τον αγώνα και για πιο λόγο, ο δημιουργός προχώρησε στο κομμάτι της προετοιμασίας των δεδομένων. Το πιο κρίσιμο κομμάτι του προβλήματος ήταν ότι τα δεδομένα δεν ήταν ισορροπημένα καθώς υπήρχαν πολλοί οδηγοί στις πρώτες θέσεις. Για την πρόβλεψη των θέσεων δημιουργήθηκαν δύο κλάσεις που αντιστοιχούν στο 1 και 0, όπου 1 σημαίνει ότι σκόραρε πόντους ενώ το 0 όχι. Για τη μοντελοποίηση του προβλήματος χρησιμοποιήθηκαν δύο διαφορετικά μοντέλα Random Forest και XGBoost. Έχοντας τα κατάλληλα δεδομένα χρησιμοποιήθηκε ο GridSearchCV ώστε να επιλεγούν διαφορετικές παράμετροι και να βρεθούν οι καλύτερες από αυτές και για τους δύο αλγορίθμους. Σαν αποτέλεσμα και οι δύο αλγόριθμοι κατέληξαν στο ίδιο σχεδόν αποτέλεσμα. Ωστόσο επειδή δεν ήταν σωστό να γίνει παραπλάνηση του αλγορίθμου, στο ότι ο οδηγός θα σκοράρει πόντους χωρίς η στρατηγική της ομάδας του να είναι καλή, για αυτό το λόγο έγινε χρήση της Precision ως την κύρια μετρική για να τιμωρήσουμε τις περιπτώσεις όπου όλοι οι οδηγοί βρίσκονται εντός της βαθμολογίας. Στην Formula 1 βαθμούς δέχονται μόνο οι οδηγοί που τερματίζουν στις 10 πρώτες θέσεις.

3.6 Το Acronis ενισχύει τις διαδικασίες της αεροδυναμικής στην Toyota Gazoo Racing ώστε να αναζητήσει τρόπους για να βελτιώσει την επίδοσή της

Μία τεράστια αυτοκινητοβιομηχανία όπως είναι η Toyota και συγκεκριμένα το τμήμα της που συμμετέχει στον σχεδιασμό των αγωνιστικών αυτοκινήτων [32] σε διάφορες μορφές αγώνων χρησιμοποίησε την εταιρία Acronis και το πρόγραμμά της για τεχνητή νοημοσύνη ώστε να εξελίξει την αεροδυναμική του αγωνιστικού της που θα συμμετείχε στον αγώνα αντοχής, στις 24 ώρες του Le Mans.

Η Toyota Gazoo Racing είναι η επωνυμία της Toyota στον μηχανοκίνητο αθλητισμό. Ένας από τους πολλούς της τομείς ασχολείται με τον σχεδιασμό και ανάπτυξη των αγωνιστικών αυτοκινήτων για το παγκόσμιο πρωτάθλημα αντοχής το οποίοι περιλαμβάνει και τις 24 ώρες του Le Mans. Η TGR δοκιμάζει τα αυτοκίνητά της σε

αεροσήραγγες χρησιμοποιώντας υπολογιστική ρευστοδυναμική το οποίο παράγει αρκετό όγκο δεδομένων. Είχαν προσπαθήσει με αρκετούς τρόπους να εκμεταλλευτούν αυτό τον όγκο δεδομένων αλλά μόνο η τεχνητή νοημοσύνη και η μηχανική μάθηση έχουν τη δυνατότητα να εκμεταλλευτούν αυτά τα δεδομένα. Οπότε έβαλαν ως στόχο να εκπαιδεύσουν έναν πράκτορα τεχνητής νοημοσύνης στο τι να αναζητήσει. Αρχικά έπρεπε να αποδειχθεί ότι η τεχνολογία του Acronis είναι λειτουργική για εργασία που τον χρειαζόντουσαν. Έτσι ανέθεσαν στο Acronis με βάση τα δεδομένα που πρόκυπταν από την υπολογιστική ρευστοδυναμική και της αεροσήραγγας να βρει βελτιώσεις στο καινούργιο υβριδικό αγωνιστικό της TGR. Από την αρχή το Acronis έδειξε τη δυναμική του καθώς βοήθησε την ομάδα της TGR να κατανοήσει την αεροδυναμική επίδοση του αγωνιστικού και να απομονώσουν τις περιοχές που θέλουν βελτίωση. Με τη χρήση του Acronis η ομάδα της TGR καταφέρνει να ξοδεύει λιγότερο χρόνο για την αναζήτηση δεδομένων και ελευθερώνεται χρόνος για αναζήτηση περισσότερων βελτιώσεων. Έτσι η ομάδα γίνεται πιο παραγωγική και μπορεί να εστιάσει εκεί που πρέπει χωρίς να σπαταλά χρόνο.

3.7 Αυτόνομο drift με μεγάλη ταχύτητα χρησιμοποιώντας ενισχυτική μάθηση

Οι Cai et al. [33] κατάφεραν τον αυτόνομο έλεγχο πλαγιολήσθσης ενός αυτοκινήτου σε εικονικό περιβάλλον με τη χρήση βαθιάς ενισχυτικής μάθησης και το οποίο μπορεί πολύ εύκολα να εφαρμοστεί και σε ένα πραγματικό περιβάλλον.

Στον κόσμο του ράλι, η πλαγιολίσθηση σε μία στροφή είναι γνωστό ως drift. Με στόχο οι επαγγελματίες οδηγοί αγώνων να διασχίσουν μία απότομη στροφή εκτελούν dirft σκοπίμως ώστε να αποσταθεροποιήσουν το όχημα τους και να διασχυθεί η στροφή χάνοντας τον λιγότερο δυνατό χρόνο. Καθώς όμως η γωνία της πλαγιολίσθησης αυξάνεται, αυξάνεται και η δυσκολία ελέγχου του οχήματος καθώς υπάρχει μεγαλύτερη αποσταθεροποίηση και μεγαλύτερος βαθμός δυσκολίας στον έλεγχό του. Το γεγονός ότι οι οδηγοί σκοπίμως κάνουν drift σημαίνει ότι υπάρχει αρκετή γνώση που πρέπει ερευνηθεί για τον έλεγχο του οχήματος όπως είναι οι γρήγορες και συχνές αλλαγές στην κίνηση του τιμονιού και ο έλεγχος της πίεσης του γκαζιού που πρέπει να εκτελεστούν με ακρίβεια ώστε να μην χαθεί ο έλεγχος

του οχήματος.

Αρχικά σχεδιάστηκε ένας μικροελεγκτής κλειστού βρόγχου σε βαθιά ενισχυτική μηχανική μάθηση χωρίς μοντέλο για τον έλεγχο προσθιοκίνητων αυτοκινήτων κατά την οδήγησή τους σε μεγάλη ταχύτητα και τον έλεγχο σε drift καθώς περνούν γρήγορα από απότομες στροφές ακολουθώντας μία προκαθορισμένη σειρά ενεργειών που έχουν θέσει οι ερευνητές. Αφού έγιναν αξιολογήσεις στο συγκερκιμένο πρόβλημα έπειτα έγινε γενικοποίηση του προβλήματος ώστε να χρησιμοποιηθούν και άλλοι τύποι οχημάτων, στροφών και διαφορετικής τριβής ελαστικών. Δημιουργήθηκαν επτά διαφορετικοί χάρτες όπου ο καθένας είχε διαφορετική δυσκολία βασισμένοι στο παιχνίδι PopKart. Σε μερικά από αυτά τα περιβάλλοντα οι προκαθορισμένες γνώσεις ήταν αρκετές για τον έλεγχο του οχήματος. Στα υπόλοιπα όμως, για την αξιολόγηση του ελεγκτή ήταν αναγκαία η χρήση ενός έμπειρου οδηγού αγώνων ο οποίος τοποθετήθηκε στο παιχνίδι με πεντάλ και τιμόνι ώστε να γίνει καταγραφή των κινήσεών του στους διαφορετικού χάρτες. Ο ελεγκτής αρχικά εκπαιδεύεται και στους έξι χάρτες ώστε να μάθει εύκολες κινήσεις, όπως είναι απλή επιτάχυνση και να εκτελεί drift σε απλές στροφές. Καθώς αυξάνεται ο βαθμός δυσκολίας ο ελεγκτής χρησιμοποιεί τα δεδομένα που έχει αποκτήσει καθώς τώρα σε κάθε επεισόδιο επιλέγεται τυχαία ο χάρτης που εφαρμόζει τις γνώσεις του. Για τη σύγκριση των δεδομένων χρησιμοποιήθηκαν τρεις διαφορετικές μεθόδους Deep Q-Network, Deep Deterministic Policy Gradient και Soft Actor-Critic, οι οποίοι δοκιμάστηκαν τέσσερις φορές σε όλους του χάρτες. Από τα αποτελέσματα προέκυψε ότι η μέθοδος SAC έχει τέλεια επίδοση στο να ακολουθεί την τροχιά των περισσότερων στροφών. Ο DQN για να διατηρήσει σταθερό το όχημα χρησιμοποιεί μικρότερες τιμές του γκαζιού ενώ ο DDPG έχει τρομακτική ταρακούνηση στην κίνηση του τιμονιού κατά τη διάρκεια του περάσματος από τη στροφή.

3.8 Η τεχνητή νοημοσύνη λαμβάνει θέση στον αγώνα

Η εταιρία Ansys δημιοσίευσε στο blog της [34], με αρθρογράφο τον Jamie Gooch, ένα άρθρο για τη θέση της τεχνητής νοημοσύνης και πως συγκεκριμένα οι μηχανικοί στο Nascar την εκμεταλλεύονται για να έχουν το καλύτερο δυνατό αποτέλεσμα στον αγώνα .

Οι σημερινές ομάδες του μηχανοκίνητου αθλητισμού βρίσκονται στο προσκήνιο

των τεχνολογικών ανακαλύψεων είτε στην προσπάθεια τους να γλιτώσουν και το παραμικρό δευτερόλεπτο μέσα στην πίστα είτε προσπαθώντας να βελτιώσουν την ακεραιότητα του οχήματος σε εκτός δρόμου διαδρομές. Πιο συγκεκριμένα η ομάδα του NASCAR η Richard Childress Racing(RCR) χρησιμοποιεί προσομοιώσεις και υπερυπολογιστές για χρόνια με σκοπό να τους βοηθήσουν να πάρουν οδηγούμενες από δεδομένα αποφάσεις πριν από κάθε αγώνα. Κατά την ημέρα ενός αγώνα η ομάδα της RCR και κάθε άλλη ομάδα λαμβάνει δεδομένα κάθε 5 δευτερόλεπτα συμπεριλαμβανόμενου της θέσης GPS, δεδομένα επιτάχυνσης και άλλες πληροφορίες. Δεν λαμβάνουν όμως μόνο τα δεδομένα από το δικό τους αυτοκίνητο αλλά και από αυτοκίνητα των υπόλοιπων ομάδων. Καθώς όλες οι ομάδες είχαν στη διάθεσή τους όλα αυτά τα δεδομένα οδηγήθηκαν στη χρήση αναλυτικών δεδομένων, ψηφιακά δίδυμα και εργαλεία τεχνητής νοημοσύνης. Με τη χρήση αυτών των εργαλείων έχουν πλέον τη δυνατότητα να λαμβάνουν σε πραγματικό χρόνο δεδομένα από το αυτοκίνητο και να δημιουργούν εφαρμόσιμες πρακτικές πάνω το όχημα τη στιγμή που εξελίσεται ο αγώνας. Εξετάζουν τις τιμές των οδηγών άλλων ομάδων όσον αφορά το πως στρίβουν, την ένταση στο φρενάρισμα και στο γκάζι ώστε να κατανοήσουν που μπορούν βελτιώσουν οι ίδιοι το όχημα τους. Για παράδειγμα κατά τη διάρκεια του αγώνα έχουν στη κατοχή τους ένα πρόγραμμα βελτιστοποίησης στρατηγικής το οποίο προβλέπει πότε είναι η κατάλληλη στιγμή για τον οδηγό να εισέλθει στα πιτ, πόσο γρήγορα φθείρονται τα λάστιχα, πως ο καιρός επηρεάζει την απόδοσής τους και άλλα. Με αυτό η ομάδα της RCR μπορεί να εκμεταλλευτεί τις δικιές της δυνατότητες αλλά και τις αδυναμίες των αντιπάλων της.

3.9 Το αγωνιστικό αυτοκίνητο του TUM που κέρδισε τον αγώνα Indy αυτόνομης δοκιμασίας

Τον Οκτώβριο του 2021 υπήρχε μία δοκιμασία με αυτόνομα Indy αγωνιστικά οχήματα όπου συμμετείχε και το Τεχνικό Πανεπιστήμιο του Μονάχου [35] όπου με επιτυχία κατέκτησαν την πρώτη θέση και απέκτησαν γνώσεις χρήσιμες για την εξέλιξη διαφόρων είδους οχημάτων. Την εμπειρία τους τη δημοσίευσαν στη σελίδα του Πανεπιστημίου τους.

Ένα Σάββατο σε έναν αγώνα στο Indianapolis τα αγωνιστικά αυτοκίνητα δεν εί-

χαν χειριστή έναν άνθρωπο αλλά ήταν αυτόνομα. Πανεπιστήμια από όλο τον κόσμο κλήθηκαν να δημιουργήσουν συστήματα βασισμένα στη τεχνητή νοημοσύνη ώστε να οδηγήσουν την πίστα αυτόνομα. Υπήρχαν υψηλές απαιτήσεις καθώς οι ενέργειες των άλλων αγωνιστικών είναι απρόσμενες και για αυτό το λόγο σε ταχύτητες που αγγίζουν τα 300 χιλιόμετρα το λογισμικό θα πρέπει να είναι ταχύτατο στις αντιδράσεις του. Το κάθε όχημα σε κλάσματα δευτερολέπτου συλλέγει δεδομένα από όλους τους αισθητήρες του οχήματος και τα χρησιμοποιεί ώστε να προβλέψει που κινούνται τα άλλα οχήματα και να λάβει αποφάσεις για τις ενέργειες που θα κάνει στο τιμόνι, το γκάζι και τα φρένα. Η ομάδα του Πανεπιστημίου του Μονάχου με σκοπό να είναι έτοιμοι για τον πραγματικό κόσμο εκτλέσαν αρχικά εικονικούς αγώνες με συνολικά οκτώ αγωνιστικά ώστε να μπορέσουν να αναγνωρίσουν και να διορθώσουν ό,τι λάθος υπήρχε δίνοντάς του έτσι πλεονέκτημα απέναντι στις αντίπαλες ομάδες την ώρα που θα έπρεπε το λογισμικό τους να εφαρμοστεί στις πραγματικές συνθήκες. Με τη νίκη τους διαγωνισμό, η ομάδα του TUM κατάφεραν να βελτιστοποιήσουν στις άμεσες αντιδράσεις του οχήματός του σε απρόσμενα συμβάντα που λάμβαναν χώρα σε αρκετά υψηλές ταχύτες, φέρνοντάς τους ένα βήμα πιο κοντά την ανάπτυξη αυτόνομων οχημάτων που κινούνται στους καθημερινούς δρόμους.

3.10 F1 και AWS

Τη δυναμική των Amazon Web Services η Formula 1 την έχει εκμεταλλευτεί σε πολλούς τομείς. Ένας από τους πολλούς, παρουσιάζεται στο blog της Amazon [36] που είναι η δημιουργία διαγραμμάτων με τη βοήθεια της μηχανικής μάθησης τα οποία ενισχύουν την εμπειρία των θεατών καθώς παρακολουθούν έναν αγώνα F1.

Οι γνώσεις που παρέχονται από την AWS αλλάζουν την εμπειρία του θεατή πριν, κατά τη διάρκεια και μετά τον αγώνα. Η F1 προσπαθεί να δώσει τη δυνατότητα στους θεατές να κατανοήσουν πως οι οδηγοί παίρνουν αποφάσεις σε κλάσματα δευτερολέπτων και πως οι ομάδες παίρνουν καθοριστικές αποφάσεις τις οποίες υιοθετούν στις στρατηγικές του αγώνα και οι οποίες έχουν καθοριστικές επιπτώσεις, καλές ή κακές, στην εξέλιξη του αγώνα. Μερικά παραδείγματα είναι δημιουργία οπτικών διαγραμμάτων όπου οι θεατές μπορούν να αναλύσουν τη δυναμική της κάθε ομάδας, χρησιμοποιώντας δεδομένα από τον ρυθμό των οδηγών και την κατάσταση της πίστα μπορούν γίνουν προβλέψεις για τις επερχόμενες μάχες που θα

γίνουν στον αγώνα. Επιπλέον με τα γραφήματα στρατηγικής οι θεατές γνωρίζουν κατά πόσο αποδοτική θα είναι η επιλογή της κάθε στρατηγικής που επιλέγει η κάθε ομάδα. Ακόμα ένα από τα διαγράμματα που έχουν πρόσβαση οι θεατές είναι της προβλεπόμενης στρατηγικής τα οποία προκύπτουν από ιστορικά δεδομένα στρατηγικών που είχε εφαρμόσει η κάθε ομάδα σε προηγούμενες σεζόν της Formula 1. Όλα αυτά τα διαγράμματα δημιουργούνται με τη χρήση επαναστατικών τεχνολογιών, με τη χρήση μηχανικής μάθησης και την απαραίτητη χρήση των υπερυπολογιστών που χωρίς αυτούς τίποτα δεν θα μπορούσε να συμβεί.

Κεφάλαιο 4

Περιγραφή του Προγράμματος Ενισχυτικής Μηχανικής Μάθησης για τη Δημιουργία Ενός Εικονικού Μηχανικού Στρατηγικής με τη Χρήση Τεχνητών Νευρωνικών Δικτύων

Τρία μέλη από το Ινστιτούτο Τεχνολογικής Αυτοκινητοβιομηχανίας του Πανεπιστημίου του Μονάχου και ένας μέλος του αγωνιστικού τμήματος της BMW, βασιζόμενοι στην έρευνα που είχε διεξάγει ένα από τρία μέλη, ο Thomaser, δημιούργησαν ένα πρόγραμμα σε γλώσσα προγραμματισμού Python για την αυτοματοποιημένη επιλογή στρατηγικής σε αγώνες του μηχανοκίνητου αθλητισμού και συγκεκριμένα στη Formula 1 [37]. Το πρόγραμμα αυτά χρησιμοποιήθηκε στα πλαίσια της διπλωματικής αυτής, όπου και επεκτάθηκε με μία νέα συνάρτηση ανταμοιβής.

4.1 Δομή του προγράμματος

Έχοντας στη διαθεσιμότητά μας αρκετά δεδομένα για τη σεζόν της Formula 1 το 2019 όπως είναι οι πραγματικές στρατηγικές που ακολούθησαν οι οδηγοί, στατιστικά για το πόσο διαρκεί ένα pitstop σε κάθε πίστα, δεδομένα για τη φθορά της κάθε γόμας ελαστικών και πολλά άλλα, μπορούμε να τα εκμεταλευτούμε αναλόγως.

Η εκμετάλλευση των δεδομένων γίνεται στο πρώτο μέρος τους προγράμματος

στο οποίο τα εξερευνούμε και εκμεταλλευόμαστε τη γνώση που αποκτάμε κατά τη διάρκεια των προσμοιώσεων ώστε να δημιουργήσουμε ένα μοντέλο δεδομένων. Το μοντέλο αυτό προκύπτει από συνεχείς προσμοιώσεις αγώνων Formula 1. Κάθε φορά που τελειώνει μία προσομοίωση κάνουμε δειγματοληψία των παρατηρήσεων που λαμβάνουμε ώστε να δημιουρήγουμε το μοντέλο στρατηγικής, το οποίο αποτελείται από τις τιμές Q-Value για τη χρήση μία από τις τέσσερις διαθέσιμες γόμες ελαστικών σε κάθε γύρο. Το μοντέλο αυτό της στρατηγικής που δημιουργείται το εκμεταλλευόματε στο δεύτερο μέρος του προγράμματος όπου δεν εξερευνούμε νέα δεδομένα αλλά χρησιμοποιούμε το μοντέλο για να προσομοιώσουμε μία διεξαγωγή αγώνα με σκοπό στο τέλος της προσομοίωσης ο εικονικός πράκτορας στρατηγικής να δημιουργήσει μέσω της ενισχυτικής μάθησης την καλύτερη δυνατή στρατηγική ώστε ο οδηγός που έχουμε επιλέξει η στρατηγική του να επιλεχθεί μέσω ενισχυτικής μάθησης και έχωντας ως αντιπάλους τους υπόλοιπους οδηγούς με τις πραγματικές του στρατηγικές, να καταχθεί στην καλύτερη δυνατή θέση.

4.2 Περιγραφή του κώδικα του προγράμματος

Στο πρόγραμμά μας επιλέγουμε αρχικά έναν οδηγό ο οποίος θα έχει έναν εικονικό μηχανικό στρατηγικής και ο οποίος θα λαμβάνει αποφάσεις με βάση την ενισχυτική μηχανική μάθηση και με αντιπάλους τους υπόλοιπους οδηγούς με τις πραγματικές τους στρατηγικές.

Στις περισσότερες μορφές μηχανοκίνητου αθλητισμού όπου τα οχήματα είναι οριοθετημένα μέσα σε μία πίστα το σημαντικότερο στοιχείο πέρα από τις ικανότητές του οδηγού και τις δυναμικές του οχήματος που οδηγεί καθένας τους είναι η στρατηγική που επιλέγουν. Σε ένα άθλημα όπως είναι η Formula 1 η σωστή επιλογή στρατηγικής είναι αρκετά σημαντική καθώς και το παραμικρό δευτερόλεπτο είναι αρκετά σημαντικό όταν ένας οδηγός στοχεύει προς την καλύτερη δυνατή θέση και τη νίκη. Αυτός είναι και ο στόχος του προγράμματός μας, η κατάταξη του οδηγού που επιλέγουμε να είναι η καλύτερη, επιλέγοντας την καλύτερη δυνατή στρατηγική. Για να το καταφέρουμε αυτό έπρεπε να λάβουμε υπόψη παράγοντες που επηρεάζουν την επιλογή των μηχανικών των ομάδων στο αν θα εκτελέσουν ένα pitstop όπως είναι οι κίτρινες σημαίες, η παρουσία αυτοκινήτου ασφαλείας και η παρουσία εικονικού αυτοκινήτου ασφαλείας.

Έχοντας στη διαθεσιμότητά μας αρκετά δεδομένα για τη σεζόν της Formula 1 το 2019 όπως είναι οι πραγματικές στρατηγικές που ακολούθησαν οι οδηγοί, στατιστικά για το πόσο διαρκεί ένα pitstop σε κάθε πίστα, δεδομένα για τη φθορά της κάθε γόμας ελαστικών και πολλά άλλα μπορούμε να εκμεταλλευτούμε αναλόγως. Η εκμετάλλευση αυτή των δεδομένων γίνεται στο πρώτο μέρος τους προγράμματος στο οποίο τα εξερευνούμε και εκμεταλλευόμαστε τις γνώσεις που αποκτάμε κατά τις διάρκεια των προσομοιώσεων ώστε να δημιουργήσουμε ένα μοντέλο δεδομένων το οποίο το εκμεταλλευόματε στο δεύτερο μέρος του προγράμματος, ώστε να δούμε αν η στρατηγική που προέκυψε από την εκμετάλλευση του μοντέλου κατατάσει τον οδηγό στην καλύτερη δυνατή θέση. Και στις δύο περιπτώσεις χρησιμοποιούμε συγκριτικά δύο συναρτήσεις ανταμοιβής μία του δημιουργού του προγράμματος και μία που προτείνουμε ως μέρος της εργασίας. Στο πρόγραμμά μας επιλέγουμε αρχικά έναν οδηγό ο οποίος θα έχει έναν εικονικό μηχανικό στρατηγικής και ο οποίος θα λαμβάνει αποφάσεις με βάση την ενισχυτική μηχανική μάθηση και με αντιπάλους τους υπόλοιπους οδηγούς με τις πραγματικές τους στρατηγικές.

Στο πρώτο μέρος του προγράμματος γίνεται η δημιουργία του μοντέλου με τα Q-Values που θα εκμεταλλευτούμε στο δεύτερο μέρος του προγράμματος. Για να δημιουργηθεί αυτό το μοντέλο πρέπει αρχικά πρέπει να γίνει η εκπαίδευσή του. Η εκπαίδευσή γίνεται με τη χρήση ενισχυτικής μάθησης και ενός Q-Network ώστε να μπορέσουμε να αντιστοιχίσουμε τις ενέργειες του πράκτορα δηλαδή την επιλογή τεσσάρων διαφορετικών ελαστικών με τις ανταμοιβές (rewards) τους ώστε στο δεύτερο μέρος να εκμεταλλευτούμε αυτές τις τιμές με σκοπό προβλέψουμε τη στρατηγική που θα φέρει τον οδηγό μας στην καλύτερη δυνατή θέση. Αρχικά πρέπει να δημιουργηθεί το περιβάλλον του πράκτορα, το οποίο ορίζεται από τα δεδομένα που έχουμε πει ότι έχουμε στη διαθεσιμότητά μας. Επιπλέον στο περιβάλλον του πράκτορα εισάγουμε και τους εξωτερικούς παράγοντες όπως την παρουσία αυτοκινήτου ασφαλείας, κίτρινης σημαίας ή εικονικού αυτοκινήτου ασφαλείας. Σε ένα πρόβλημα ενισχυτικής μάθησης η τακτική (policy) είναι ένας παράγοντας που πρέπει να οριστεί ώστε ο πράκτορας να γνωρίζει ποια ενέργεια πρέπει να εκτελέσει σε κάθε επεισόδιο του προβλήματος. Η τακτική στο πρόγραμμά μας επιλέγει μία τυχαία ενέργεια σε κάθε επεισόδιο, δηλαδή σε κάθε γύρου του αγώνα εκτελεί μία τυχαία επιλογή ελαστικού. Έπειτα ορίζουμε το replay buffer απαραίτητο για την εκ-

παίδευση του μοντέλου, όπως δημιουργούμε και τη συνάρτηση για τη συλλογή των δεδομένων σε κάθε επεισόδιο όπου μέσα σε αυτή καλούνται οι συναρτήσεις για την παρατήρηση και η συνάρτηση reward. Στην εκπαίδευση του μοντέλου το πιο σημαντικό κομμάτι είναι τα δεδομένα που λαμβάνουμε από τις συναρτήσεις παρατήρησης και ανταμοιβής. Μέσα από τη συνάρτηση παρατήρησης ο πράκτοράς μας βλέπει την κατάσταση του περιβάλλοντος ώστε να λάβει κατάλληλες αποφάσεις. Οι αποφάσεις που λαμβάνει ο πράκτορας γίνονται κατά τη διάρκεια προσομοίωσης κάθε γύρου μέχρι το τέλος προσομοίωσης του αγώνα. Ένας αγώνας προσομοιώνεται 250 χιλιάδες φορές, όπου σε κάθε γύρο της προσομοίωσης, δηλαδή σε κάθε επεισόδιο, λαμβάνουμε τις τιμές της συνάρτησης ανταμοιβής (reward) ώστε στο τέλος της προσομοίωσης όλου του αγώνα και παίρνοντας τον μέσο όρο των τιμών ανταμοιβής να σχηματιστεί ο πίνακας με τα Q-Values.Υπάρχουν δύο συναρτήσεις reward μία των δημιουργών του προγράμματος και μία που υλοποιήθηκε στα πλαίσια της διπλωματικής εργασίας με στόχο τη σύγκριση των αποτελεσμάτων τους.

Όσον αφορά τη συνάρτηση ανταμοιβής του δημιουργού του προγράμματος αρχικά υπολογίζουμε τον συνολικό χρόνο που κάνει ο οδηγός σε ένα γύρο όταν εισέρχεται στα pits για αλλαγή ελαστικών. Έπειτα διακρίνουμε αν υπάρχει αυτοκίνητο ασφαλείας ή εικονικό αυτοκίνητο ασφαλείας ή τίποτα από τα δύο. Και στις δύο περιπτώσεις επιστρέφεται μία διαφορετική τιμή ανταμοιβής. Ωστόσο αν οδηγός και πριν το pitstop αλλά και μετά το pitstop βρίσκεται πίσω από το αυτοκίνητο ασφαλείας έχει reward ίσο με μηδέν καθώς δεν κερδίζει κάποιο πλεονέκτημα σε αυτή την περίπτωση.

Κατά τη διάρκεια την προσομοίωσης βρίσκουμε τις απώλειες που έχει η εκπαίδευση του πράκτορα και κάνουμε αξιολόγηση της εκπαίδευσης βρίσκοντας τη μέση τιμή των ανταμοιβών από αυτές τις προσομοιώσεις. Επόμενη δουλειά του κυρίως προγράμματος αφού συλλεχθούν τα δεδομένα είναι η μετατροπή των δεδομένων που έχει το Q-Network τα οποία είναι οι μέσοι όροι ανταμοιβής χρήσης, των τριών ελαστικών σε κάθε γύρο, σε ένα tf-lite μοντέλο ώστε να μπορέσουμε να εκμεταλλευτούμε αργότερα για να κατατάξουμε τον οδηγό που επιλέγουμε στην καλύτερη δυνατή θέση.

Στη συνέχεια γίνεται προσομοίωση για το τέλος του γύρου όπου υπολογίζουμε τους τελικού χρόνους για τον γύρο για όλους τους οδηγούς, αυξάνουμε τον γύρο

κατά 1 ώστε να πάμε στον επόμενο, αυξάνουμε τη συνολική φθορά όλου του αυτοκινήτου, ελέγχουμε αν κάποιος έχει αποσυρθεί από τον αγώνα, αποθηκεύουμε τους δείκτες των οδηγών που εκτέλεσαν pit-stop και τέλος γίνονται έλεγχοι για τα αποτελέσματα της προσομοίωσης του γύρου όπως είναι η χρήση δύο διαφορετικών ελαστικών όσο αναφορά τη γόμα τους και επιστρέφουμε στον πράκτορα την παρατήρηση και την ανταμοιβή για τις ενέργειές του. Αν έχουμε συμπληρώσει τον αριθμό των προσομοιώσεων συλλέγουμε τα παραπάνω δεδομένα και τερματίζει το πρόγραμμα.

Στο δεύτερο μέρος του προγράμματος χρησιμοποιούμε το μοντέλο που δημιουργήθηκε πριν ώστε να κάνουμε προσομοίωση του αγώνα και με τη βοήθεια της ενισχυτικής μάθησης να γίνει κατάλληλη επιλογής στρατηγικής ώστε ο οδηγός μας να καταταχθεί στην καλύτερη δυνατή θέση. Η διαφορά σε αυτή την περίπτωση είναι ότι εκμεταλλευόμαστε τις τιμές Q-value από το μοντέλο ώστε να αποφασίσουμε για την επιλογή μίας γόμας, αλλά δεν εξερευνούμε σε κάθε γύρο ποιο θα είναι το αποτέλεσμα επιλέγοντας κάθε φορά μία τυχαία γόμα ελαστικού. Αρχικά έχουμε ως δεδομένο πόσοι πράκτορες και πόσες φορές θα προσομοιωθεί ένας αγώνας. Στην περίπτωση μας ο πράκτορας είναι ένας και η προσομοίωση του αγώνα είναι μία. Ελέγχουμε αν τα δεδομένα στο τέλος της προσομοίωσης του αγώνα είναι έγκυρα ώστε αν είναι, να οδηγηθούμε στην εκτύπωση των αποτελεσμάτων και αν δεν είναι να επαναλάβουμε ξανά την προσομοίωση. Η προσομοίωση του αγώνα γίνεται όπως και στο πρώτο μέρος του προγράμματος, απλώς εδώ δεν τρέχουμε έναν αγώνα 250 χιλιάδες φορές, αλλά μία φορά ώστε να βγάλουμε τα αποτελέσματά του. Για τη σύγκριση των δύο συναρτήσεων ανταμοιβής, τρέχουμε το συγκεκριμένο πρόγραμμα και για τα δύο μοντέλα που έχουν δημιουργήσει οι συναρτήσεις reward και συγκρίνουμε σε ποια θέση έχει καταλήξει ο οδηγός που επιλέγουμε και στις δύο περιπτώσεις. Τα αποτελέσματα που προβάλλονται είναι σε δύο μορφές μία είναι τα τυπωμένα δεδομένα και η άλλη είναι τα γραφήματα όπου μπορούμε να δούμε συγκριτικά γραφήματα για την πορεία όλων των οδηγών στον αγώνα. Στα τυπωμένα μπορούμε να δούμε την τελική κατάταξη όλων οδηγών στο τέλος της προσομοίωσης που είναι και το σημαντικότερο για τον στόχο που έχουμε.

4.3 Η δεύτερη συνάρτηση reward στην οποία εργαστήκαμε

Στη συνάρτηση reward που προτείνουμε περιλαμβάνονται όλοι έλεγχοι που έχει ο δημιουργός του προγράμματος και ότι υπολογισμούς γίνονται στη συνάρτησή του. Αρχικά ελέγχουμε αν είναι δυνατή η επίτευξη undercut δηλαδή ο οδηγός που θα επιλέξουμε να κάνει pit-stop, να καταφέρει να προσπεράσει τον προπορευόμενό του όταν θα εκτελέσει αυτός το pit-stop του. Αν έχει τη δυνατότητα αυτή ο οδηγός μας τότε τον επιβραβεύουμε στο τελικό reward του. Στη συνέχεια γίνεται έλεγχος μετά την αλλαγή ελαστικών αν η γόμα που έχει επιλεχθεί είναι πιο γρήγορη σε χρόνο στο γύρο, όπως είναι η μαλακή. Αν είναι καλύτερη τότε αφαιρούμε μία συγκεκριμένη τιμή στο χρόνο του οδηγού στον γύρο. Αν όμως είναι χειρότερη η επιλογή που έκανε ο πράκτορας τότε προσθέτουμε μία συγκεκριμένη τιμή από τον χρόνο στον γύρο του οδηγού. Οι τιμές που προσθέτουμε ή αφαιρούμε είναι βασισμένες στις τιμές που δίνει η Pirelli για το πόσος χρόνος χάνεται ή κερδίζεται από την επιλογή μία συγκεκριμένης γόμας. Έπειτα ελέγχουμε αν τα ελαστικά που επιλέχθηκαν μπορούν να διαρκέσουν για το υπόλοιπο του αγώνα καθώς αναλόγως τη φθορά και το ελαστικό μπορεί να φθείρεται πιο γρήγορα και να μην διαρκέσει έως το τέλος του αγώνα. Οι έλεγχοι αυτοί γίνονται και στην περίπτωση που υπάρχει αυτοκίνητο ασφαλείας και στην περίπτωση που δεν υπάρχει στην πίστα. Ακόμα ελέγχουμε αν κατά τη διάρκεια του αυτοκινήτου ασφαλείας έχει επιλέξει ο πράκτορας να πραγματοποιηθεί ένα pit-stop, καθώς υπό καθεστώς αυτοκινήτου ασφαλείας χάνεται αρκετά λιγότερος χρόνος για μία αλλαγή από ότι στην κανονική ροή του αγώνα. Αν δεν έχει κάνει αυτή την επιλογή τότε τιμωρούμε τον πράκτορα στο συνολικό του reward. Τέλος αφού πραγματοποιηθούν όλοι αυτοί οι έλεγχοι επιστρέφουμε την τιμή reward στο κύριο πρόγραμμα ώστε να συμπεριληψθεί στον υπολογισμό των μέσων όρων ανταμοιβών που αποθηκεύονται στα Q-Values.

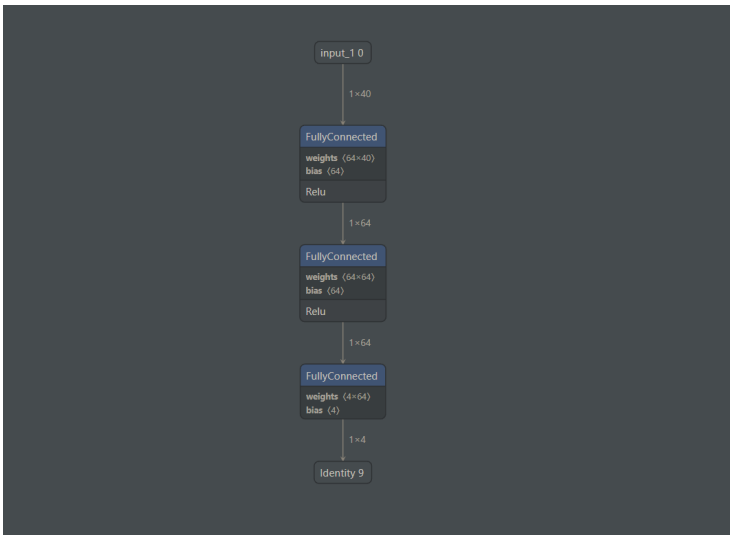
Κεφάλαιο 5

Υπολογιστική Μελέτη

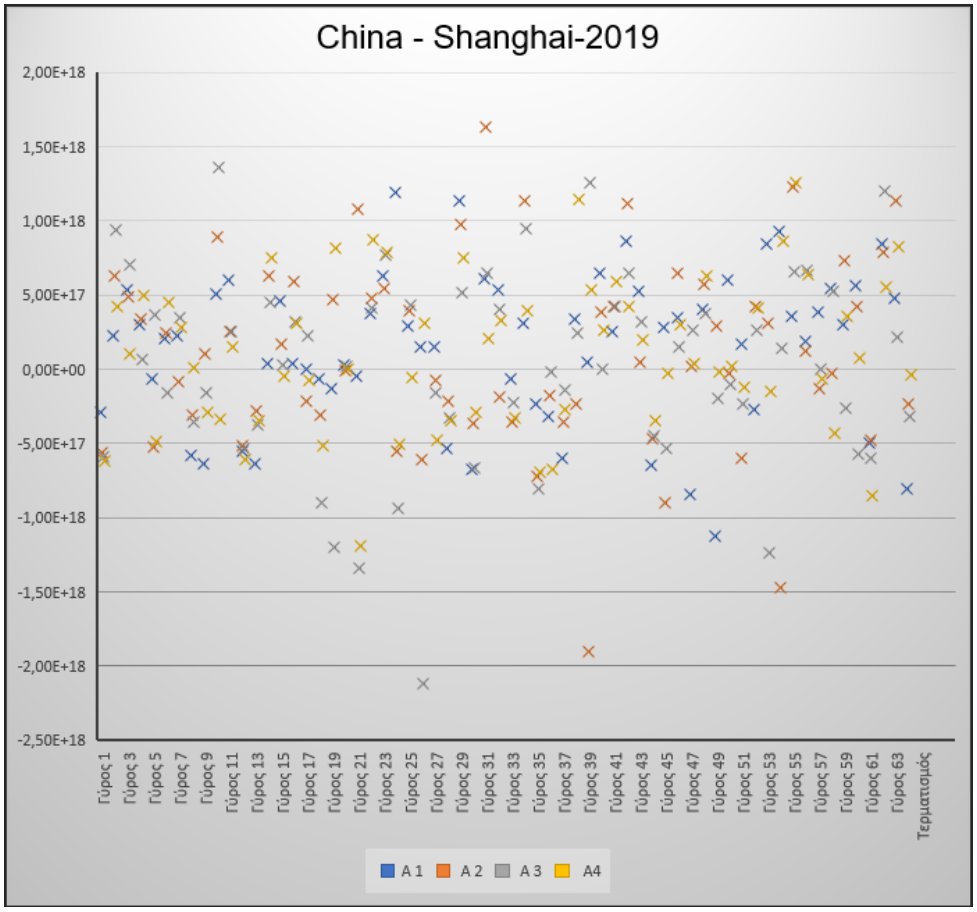
Η υπολογιστική μελέτη που διεξήγαμε περιλαμβάνει τη σύγκριση των αποτελεσμάτων που προκύπτουν από το δεύτερο μέρος του προγράμματος στο οποίο χρησιμοποιούμε ξεχωριστά κάθε φορά το μοντέλο που προέκυψε από την εκπαίδευση του πράκτορα με τη συνάρτηση ανταμοιβής των δημιουργών του προγράμματος και από την προτεινόμενη συνάρτηση ανταμοιβής. Τα μοντέλα που χρησιμοποιούμε είναι αποτελέσματα προσομοίωσης δέκα αγώνων που διεξήχθησαν την περίοδο του 2019 στη Formula 1. Σε κάθε προσομοίωση αγώνα αποφασίζαμε εμείς για ποιον οδηγό θα σχηματίσει ο πράκτορας το μοντέλο του ενάντια στους υπόλοιπους οδηγούς οι οποίοι εφαρμόζουν στην προσομοίωση την πραγματική τους στρατηγική. Δηλαδή ο κάθε οδηγός έχει το δικό του μοντέλο ενισχυτικής μάθησης και για τους 10 αγώνες στου οποίους έγινε η προσομοίωση. Κάθε οδηγός προσομοιώθηκε από 10 φορές σε κάθε έναν από τους 10 αγώνες που πραγματοποιήθηκαν την περιόδο 2019 της Formula 1. Οι προσομοιώσεις αυτές έγιναν και για τις δύο συναρτήσεις ανταμοιβής.

Το μοντέλο που δημιουργείται μετά την εκπαίδευση του πράκτορα περιλαμβάνει το νευρωνικό δίκτυο που έχει αντιστοιχίσει κάθε γόμα ελαστικού με το μέσο όρο των ανταμοιβών που προέκυψαν από την επιλογής της σε σε κάθε γύρο ξεχωριστά. Οι διαφορετικές γόμες, είναι διαφορετικές ενέργειες που μπορεί να εκτελέσει ο πράκτορας και βρίσκονται στην είσοδο του μοντέλου. Συγκεκριμένα, στο Σχήμα 5.1 μπορούμε να δούμε την είσοδό του και τα στρώματα του νευρωνικού δικτύου και το μέγεθός του. Η δομή του μοντέλου με το νευρωνικό δίκτυο φαίνεται στο Σχήμα 5.1 και οι τιμές από το μέσο όρο των ανταμοιβών κάθε γόμας σε κάθε γύρο φαίνεται στο Σχήμα 5.2.

Σχήμα 5.1: Δομή του μοντέλου



Σχήμα 5.2: Τιμές των πιθανοτήτων



Τα μοντέλα που χρησιμοποιούνται για κάθε πράκτορα τα χρησιμοποιούμε στο δεύτερο μέρος του προγράμματος, όπου αφού τα εκμεταλλευτούμε μπορούμε να συγκρίνουμε τα αποτελέσματά τους. Το κύριο μέτρο σύγκρισης που χρησιμοποιήσαμε για τις δύο συναρτήσεις είναι η κατάταξη του κάθε οδηγού στο τέλος του κάθε

αγώνα που προσομοιώνουμε. Στο Σχήμα 5.3 μπορούμε να δούμε τα αποτελέσματα που προκύπτουν από την εκμετάλλευση ενός μοντέλου του οδηγού. Μπορούμε να δούμε τη θέση κατάταξης του, τη στρατηγική που ακολούθησε ο κάθε οδηγός, τον καλύτερο του χρόνο, αν τερμάτισε τον αγώνα και τη διαφορά του από τους υπόλοιπους οδηγούς.

Σχήμα 5.3: Αποτελέσματα προσομοίωσης ενός αγώνα

```
RESULT: Simulation result:
     pos  carno   t_race    gap    int  best_t_lap  no_laps status                                  strategy_info
BOT    1     77 5464.772  0.000  0.000      94.807       56      F              [0, 'A4'], [21, 'A3'], [36, 'A4']
HAM    2     44 5480.688 15.915 15.915      94.824       56      F              [0, 'A4'], [22, 'A3'], [36, 'A4']
VER    3     33 5497.441 32.669 16.754      95.136       56      F              [0, 'A4'], [17, 'A3'], [34, 'A4']
LEC    4     16 5512.341 47.569 14.900      94.469       56      F              [0, 'A4'], [22, 'A3'], [42, 'A4']
VET    5      5 5520.370 55.597  8.029      95.688       56      F              [0, 'A4'], [14, 'A4'], [31, 'A3']
RIC    6      3 5527.058 62.286  6.688      96.599       56      F                           [0, 'A6'], [18, 'A3']
ALB    7     23 5546.099 81.326 19.041      97.028       56      F                           [0, 'A6'], [19, 'A3']
GAS    8     10 5471.880     1l     1l      95.173       55      F  [0, 'A6'], [19, 'A3'], [39, 'A4'], [53, 'A6']
MAG    9     20 5478.647     1l  6.767      96.282       55      F              [0, 'A6'], [9, 'A3'], [33, 'A4']
GRO   10      8 5479.147     1l  0.500      96.020       55      F              [0, 'A6'], [8, 'A3'], [35, 'A4']
PER   11     11 5479.647     1l  0.500      97.479       55      F                           [0, 'A4'], [20, 'A3']
SAI   12     55 5486.933     1l  7.286      96.984       55      F              [0, 'A4'], [1, 'A3'], [36, 'A6']
STR   13     18 5490.207     1l  3.274      96.041       55      F              [0, 'A4'], [20, 'A3'], [44, 'A6']
RAI   14      7 5496.770     1l  6.563      96.975       55      F                           [0, 'A4'], [25, 'A3']
GIO   15     99 5531.978     1l 35.208      97.367       55      F              [0, 'A6'], [7, 'A4'], [30, 'A4']
RUS   16     63 5548.984     1l 17.006      97.582       55      F              [0, 'A4'], [22, 'A3'], [49, 'A6']
KUB   17     88 5467.129     2l     1l      98.599       54      F                           [0, 'A4'], [26, 'A3']
NOR   18      4 4901.925     7l     5l      97.294       49    DNF   [0, 'A4'], [1, 'A3'], [17, 'A4'], [34, 'A6']
KVY   19     26 4054.999    16l     9l      96.638       40    DNF   [0, 'A4'], [7, 'A4'], [25, 'A3'], [30, 'A6']
HUL   20     27 1629.715    40l    24l      98.033       16    DNF                           [0, 'A6'], [11, 'A3']
RESULT: FCY phases: [VSC -> 82.337s - 157.360s]
RESULT: Retirements: [HUL -> 1631.580s] [KVY -> 4084.112s] [NOR -> 4941.197s]
```

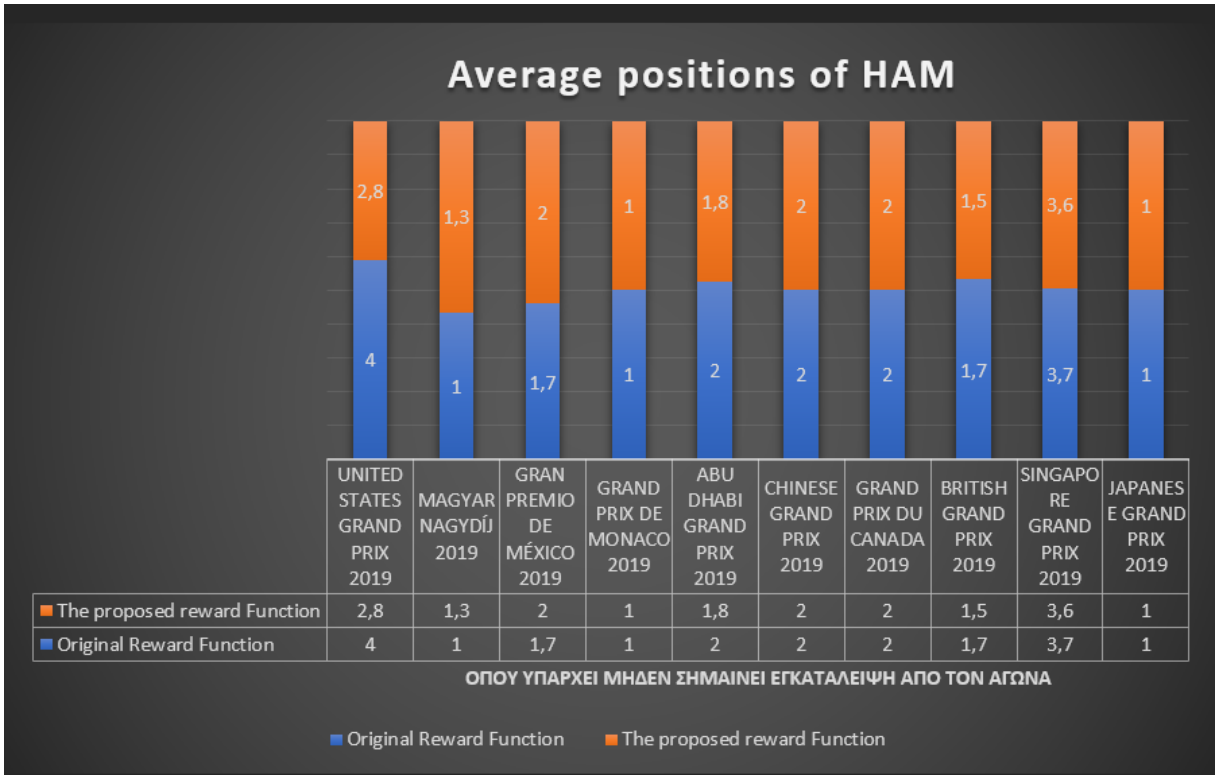
5.1 Σύγκριση δεδομένων

Για κάθε μία από τις δέκα προσομοιώσεις του κάθε αγώνα αποθηκεύσαμε τη θέση που είχε ο οδηγός στην κατάταξη στο τέλος της προσοσμοιώσης του αγώνα με σκοπό να υπολογίσουμε τον μέσο όρο της θέσης του και για κάθε μία από τις δύο συναρτήσεις reward. Και στις δύο περιπτώσεις των συναρτήσεων δεν είναι δυνατή η βελτίωση του αποτελέσματος ενός οδηγού ο οποίος παραιτείται από τον αγώνα. Επιπλέον, κάθε οδηγός έχει διαφορετικές δυνατότητες σε κάθε πίστα καθώς μπορεί να ταιριάζει το στυλ οδήγησής του και στο στήσιμο του μονοθεσίου του.

5.1.1 Αποτελέσματα προσομοιώσεων Lewis Hamilton

Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Hamilton θεωρείται ο καλύτερος και με το καλύτερο μονοθέσιο την περίοδο 2019 της Formula1. Στο Σχήμα 5.4 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Hamilton μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της

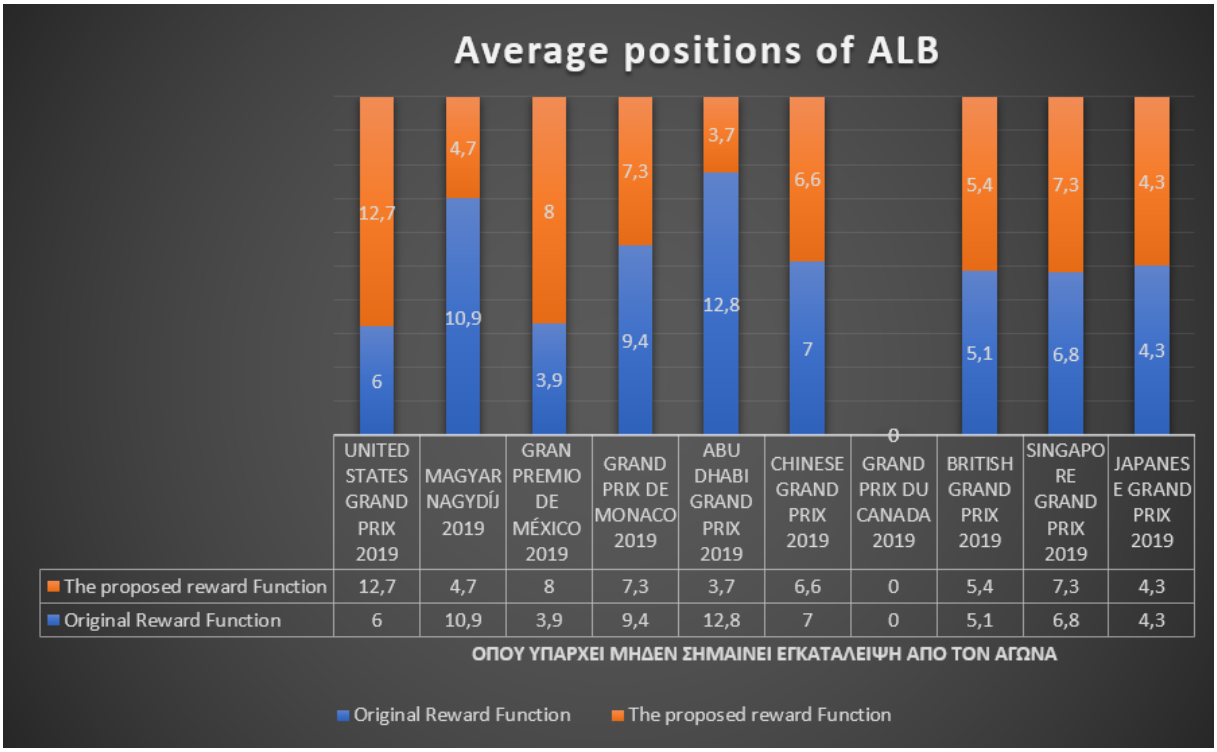
Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.4 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.4: Μέσοι όροι Lewis Hamilton

5.1.2 Αποτελέσματα προσομοιώσεων Alexander Albon

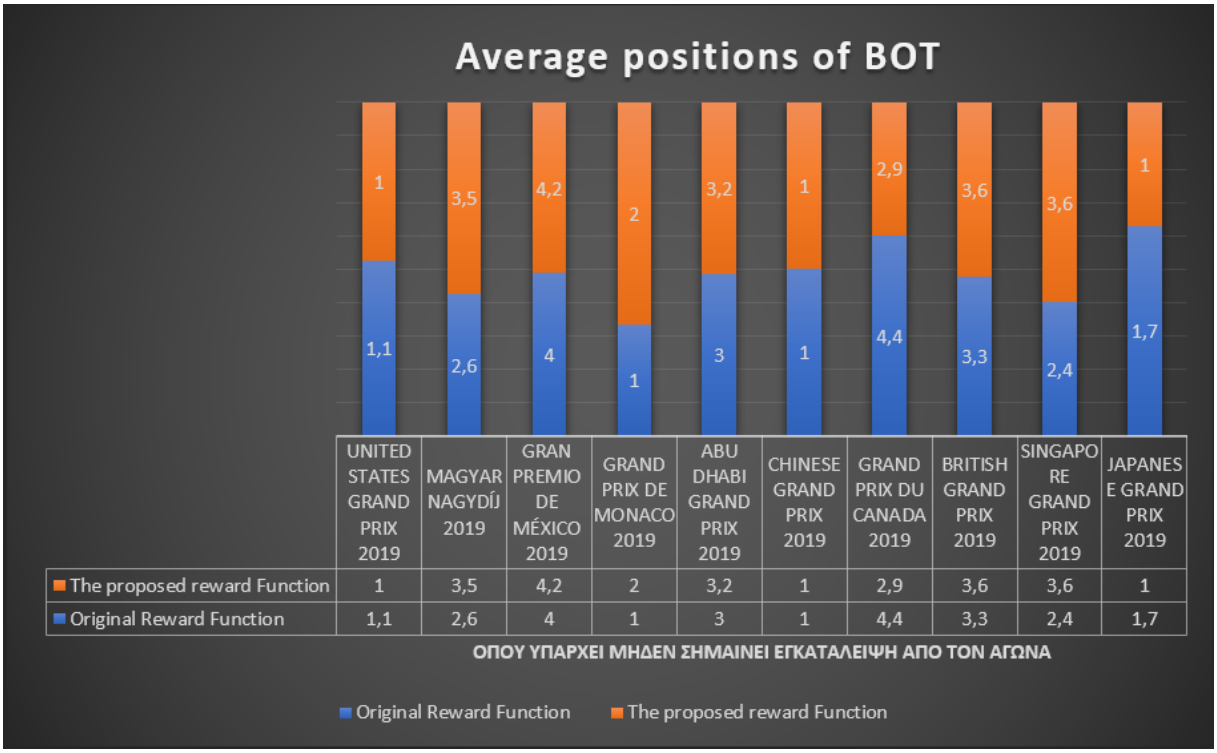
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Albon είχε αρκετά καλό μονοθέσιο, ωστόσο οι ικανότητές του ήταν μέτριες την περίοδο 2019 της Formula1. Στο Σχήμα 5.5 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Albon μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.5 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.5: Μέσοι όροι Alexander Albon

5.1.3 Αποτελέσματα προσομοιώσεων Valtteri Bottas

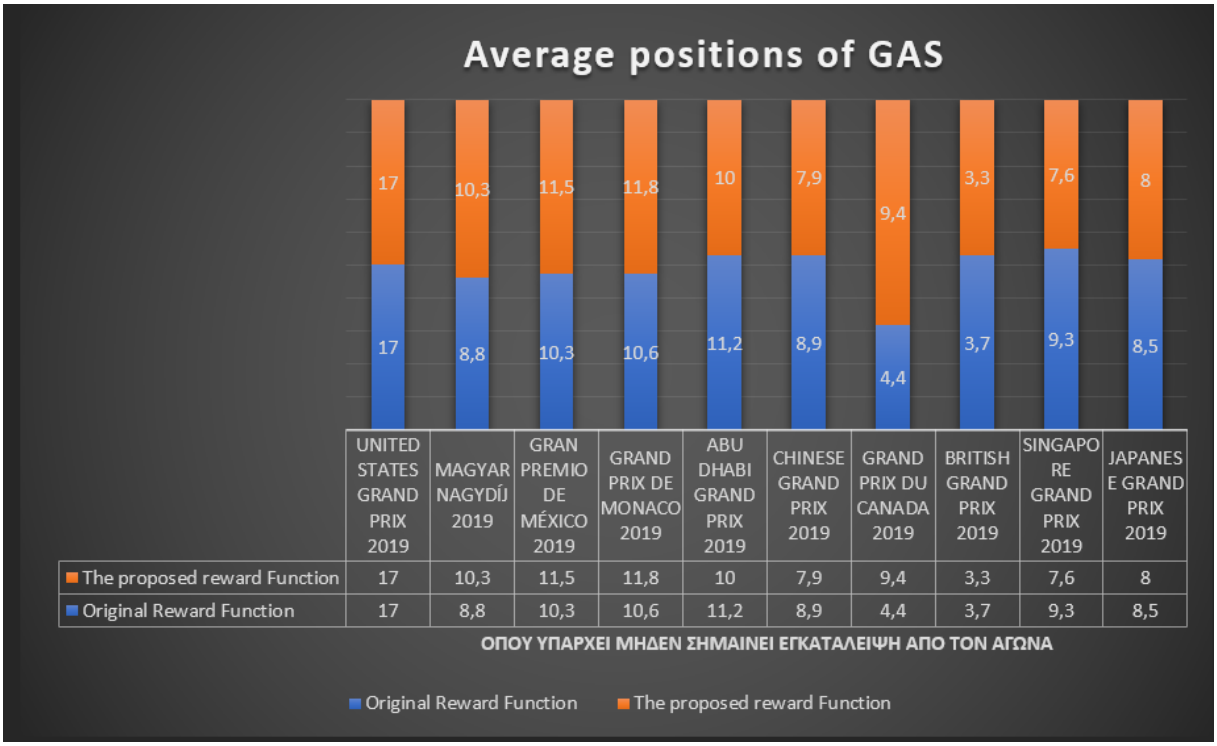
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Bottas είχε διαθέσιμο ένα από τα καλύτερα μονοθέσια στη διάθεσή, ωστόσο οι ικανότητές ήταν μέτριες την περίοδο 2019 της Formula1. Στο Σχήμα 5.6 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Bottas μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.6 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.6: Μέσοι όροι Valtteri Bottas

5.1.4 Αποτελέσματα προσομοιώσεων Pierre Gasly

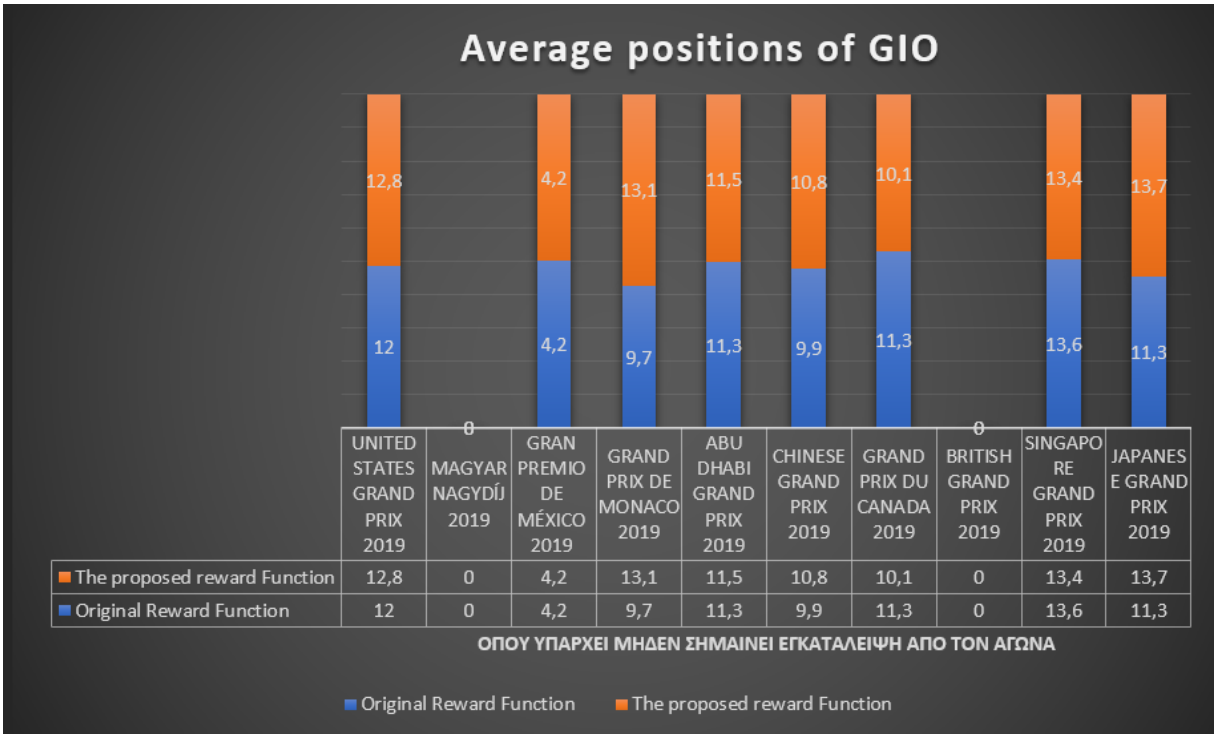
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Gasly είναι ένας αρκετά ικανός οδηγός και είχε στη διάθεσή του ένα αρκετά καλό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.7 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Gasly μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.7 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.7: Μέσοι όροι Pierre Gasly

5.1.5 Αποτελέσματα προσομοιώσεων Antonio Giovinazzi

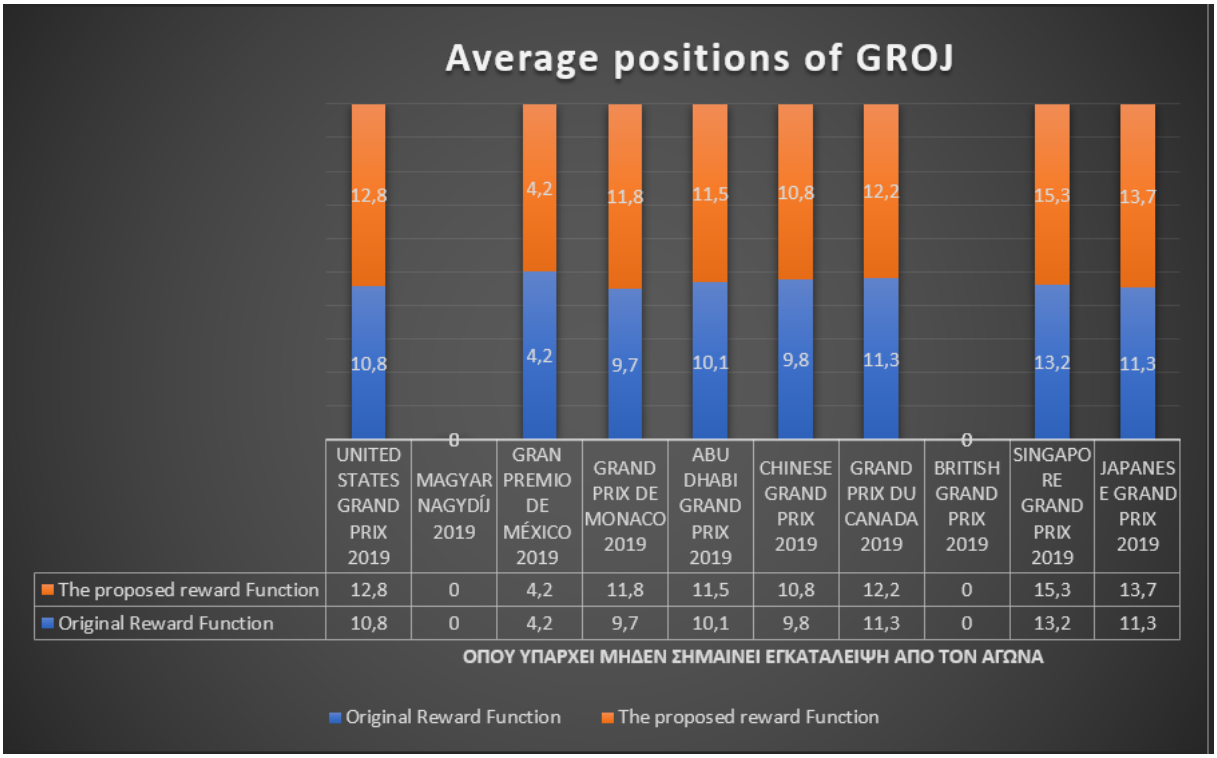
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Giovinazzi είναι ένας αρκετά ικανός οδηγός και είχε στη διάθεσή του ένα ικανό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.8 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Giovinazzi μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.8 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.8: Μέσοι όροι Antonio Giovinazzi

5.1.6 Αποτελέσματα προσομοιώσεων Romain Grojean

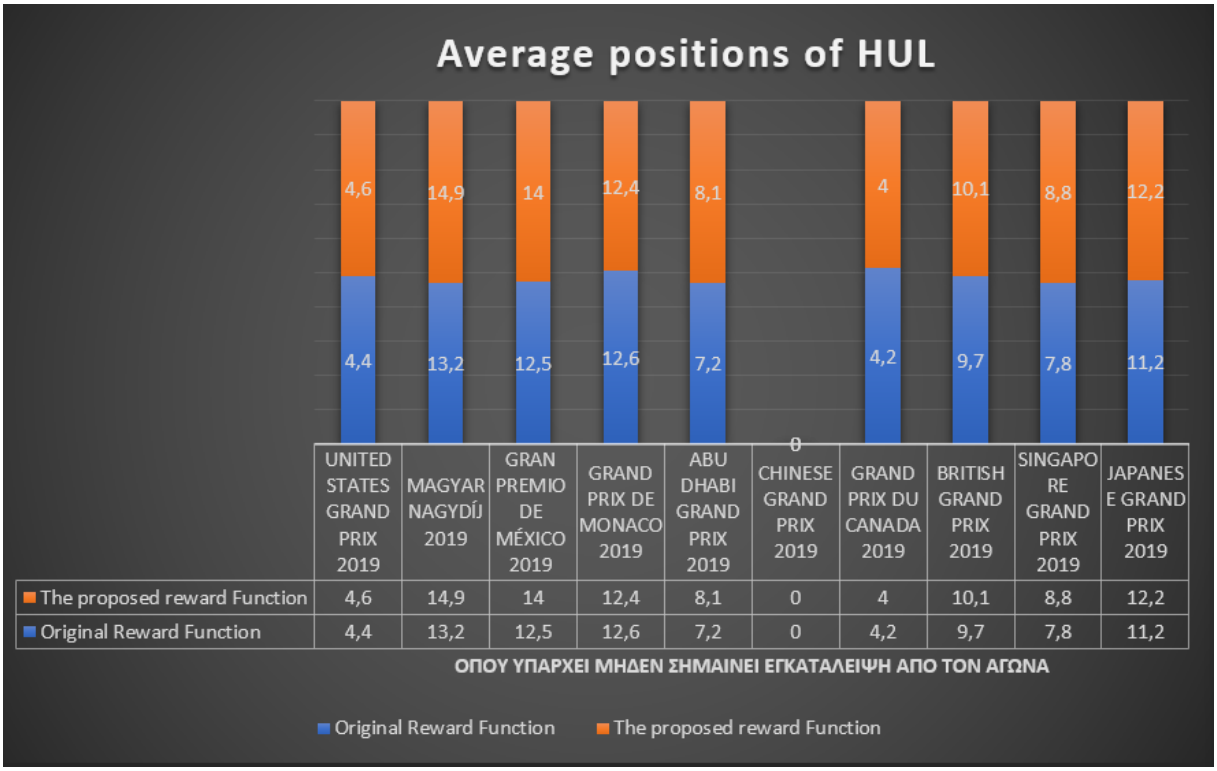
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Grojean είναι ένας μέτριος οδηγός και είχε στη διάθεσή του ένα μέτριο μονοθέσιο την περίοδο 2019 της Formula1. Στο Σχήμα 5.9 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Grojean μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.9 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.9: Μέσοι όροι Romain Grojean

5.1.7 Αποτελέσματα προσομοιώσεων Nico Hulkenberg

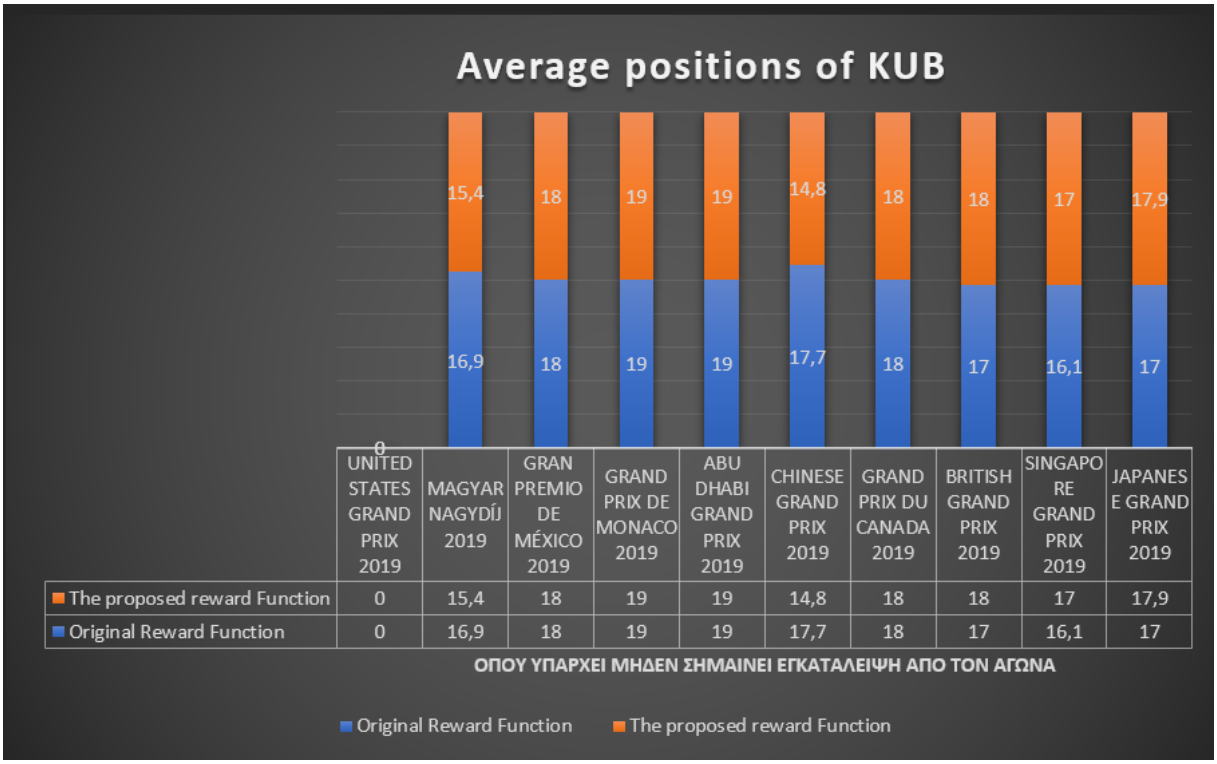
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Hulkenberg είναι ένας ένας αρκετά καλός και είχε στη διάθεσή του ένα πολύ κακό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.10 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Hulkenberg μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.10 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.10: Μέσοι όροι Nico Hulkenberg

5.1.8 Αποτελέσματα προσομοιώσεων Robert Kubica

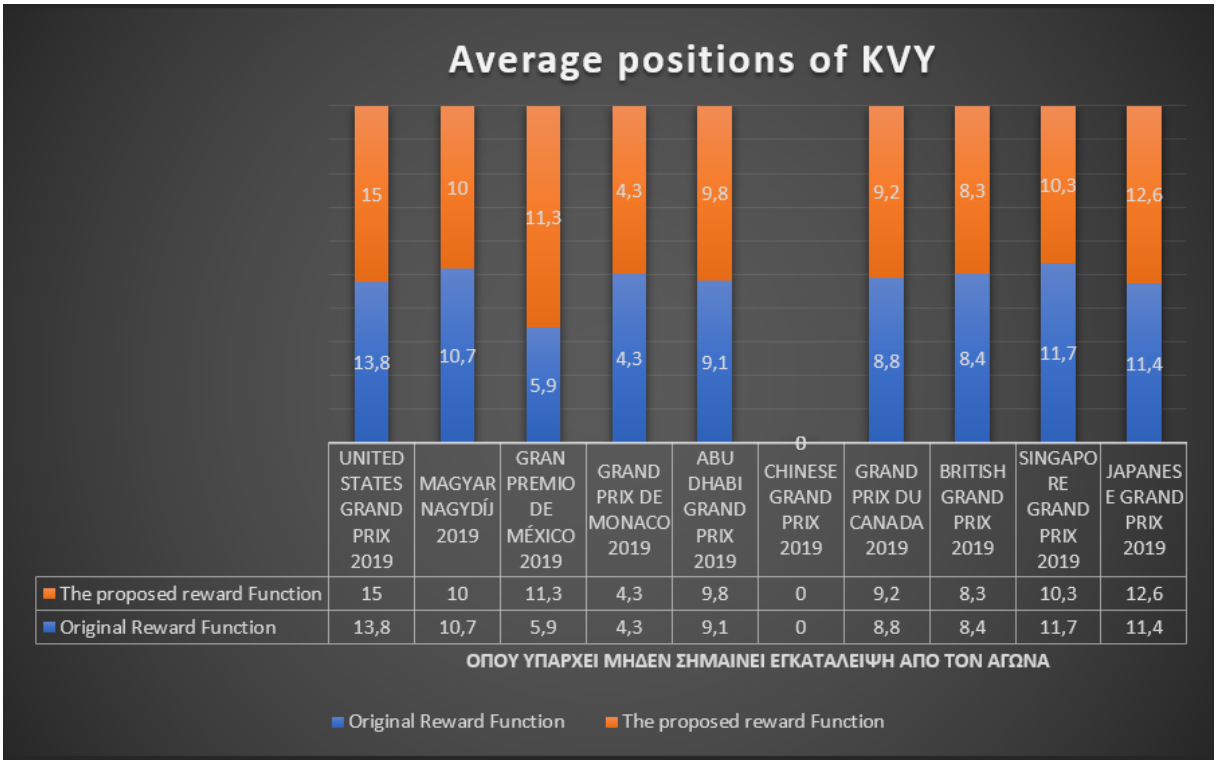
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Kubica είναι ένας μέτριος οδηγός και είχε στη διάθεσή του ένα πολύ κακό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.11 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Kubica μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.11 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.11: Μέσοι όροι Robert Kubica

5.1.9 Αποτελέσματα προσομοιώσεων Daniil Kvyat

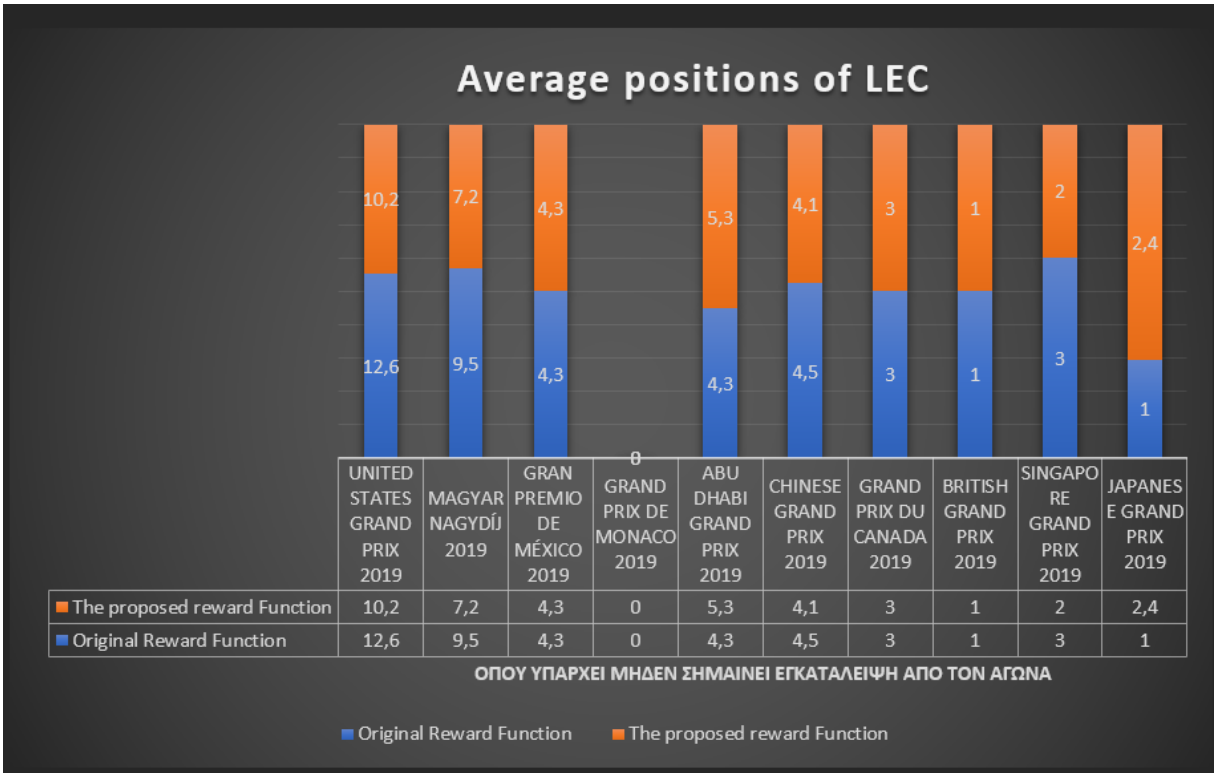
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Kvyat είναι ένας μέτριος οδηγός και είχε στη διάθεσή του ένα ικανό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.12 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Kvyat μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.12 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.12: Μέσοι όροι Daniil Kvyat

5.1.10 Αποτελέσματα προσομοιώσεων Charles Leclerc

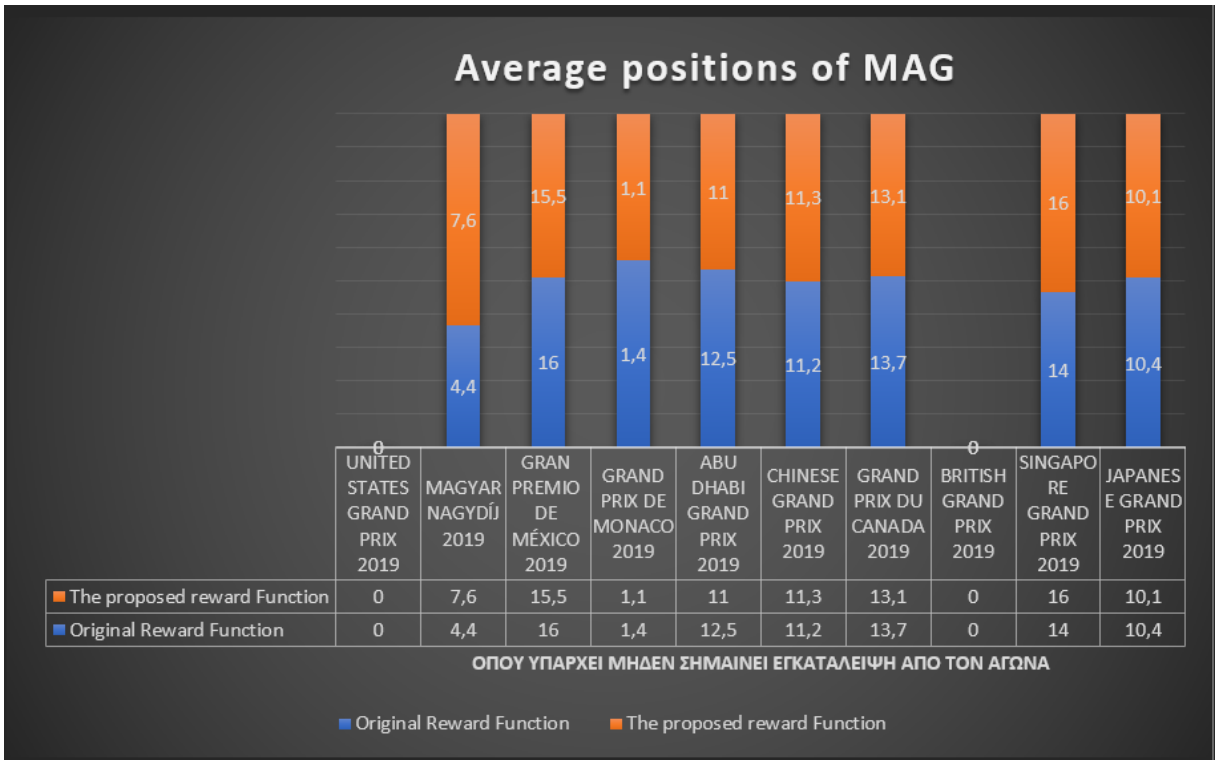
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Leclerc είναι ένας αρκετά καλός και είχε στη διάθεσή του ένα ικανό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.13 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Leclerc μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.13 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.13: Μέσοι όροι Charles Leclerc

5.1.11 Αποτελέσματα προσομοιώσεων Kevin Magnussen

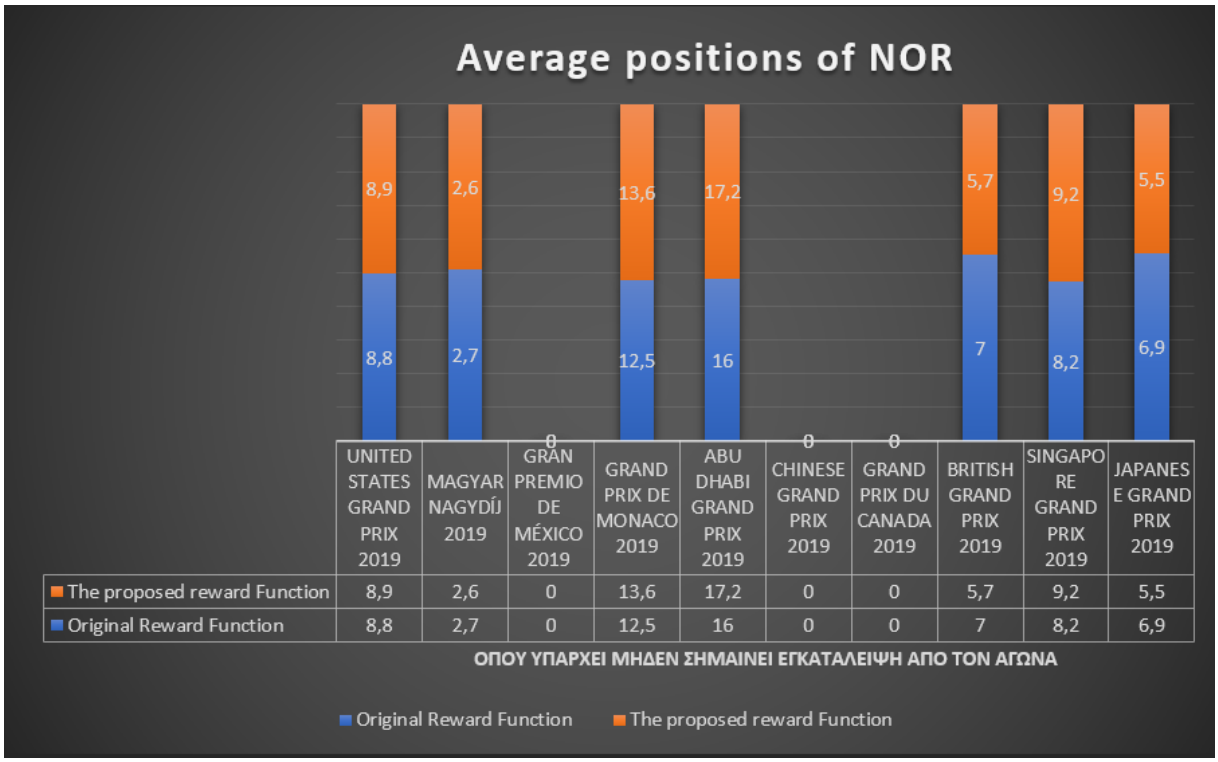
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Magnussen είναι ένας μέτριος οδηγός και είχε στη διάθεσή του ένα ικανό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.14 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Magnussen μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.14 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.14: Μέσοι όροι Kevin Magnussen

5.1.12 Αποτελέσματα προσομοιώσεων Lando Norris

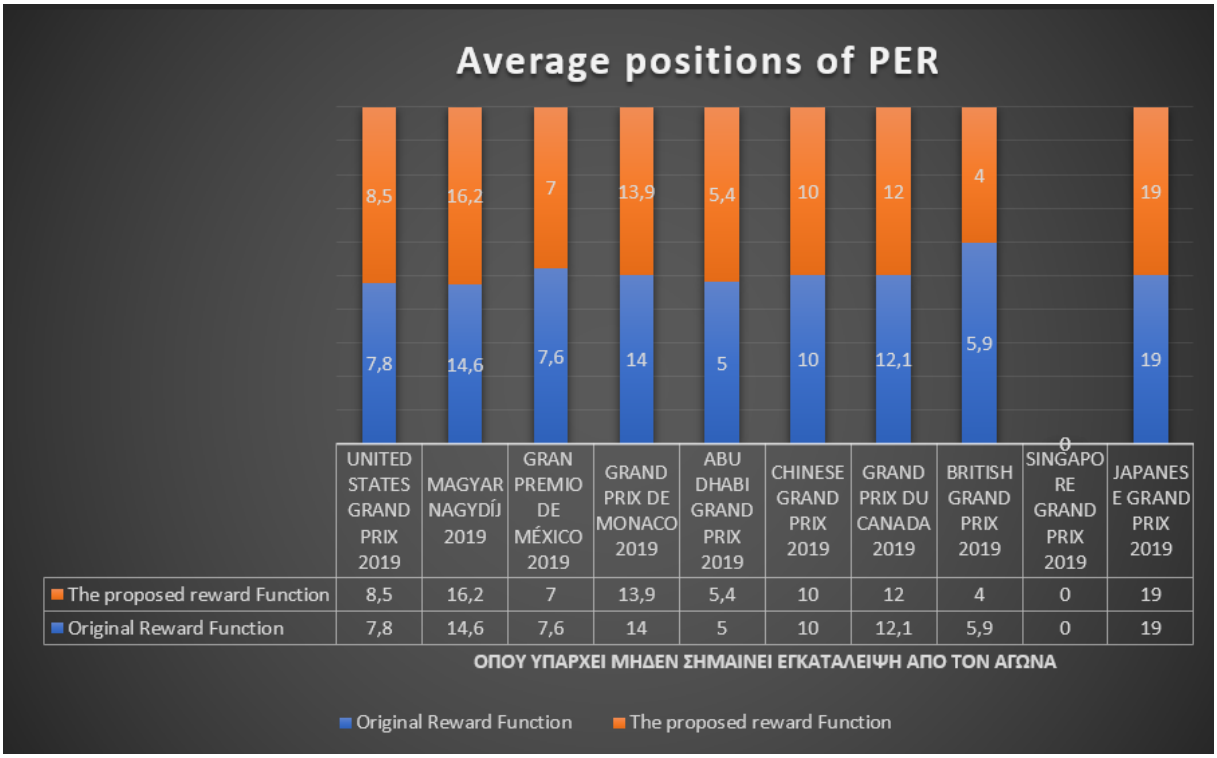
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Norris είναι ένας καλός οδηγός και είχε στη διάθεσή του ένα μέτριο μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.15 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Norris μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.15 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.15: Μέσοι όροι Lando Norris

5.1.13 Αποτελέσματα προσομοιώσεων Sergio Perez

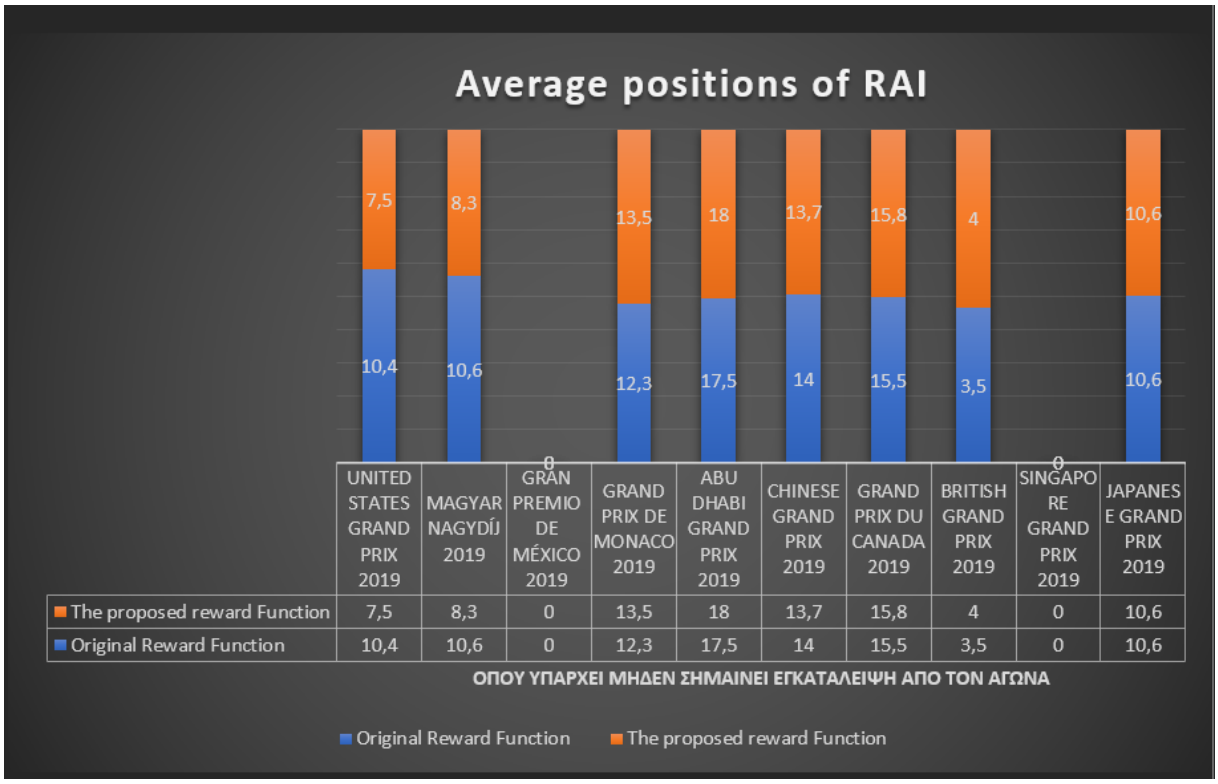
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Perez είναι ένας καλός οδηγός και είχε στη διάθεσή του ένα ικανό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.16 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Perez μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.16 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.16: Μέσοι όροι Sergio Perez

5.1.14 Αποτελέσματα προσομοιώσεων Kimi Räikkönen

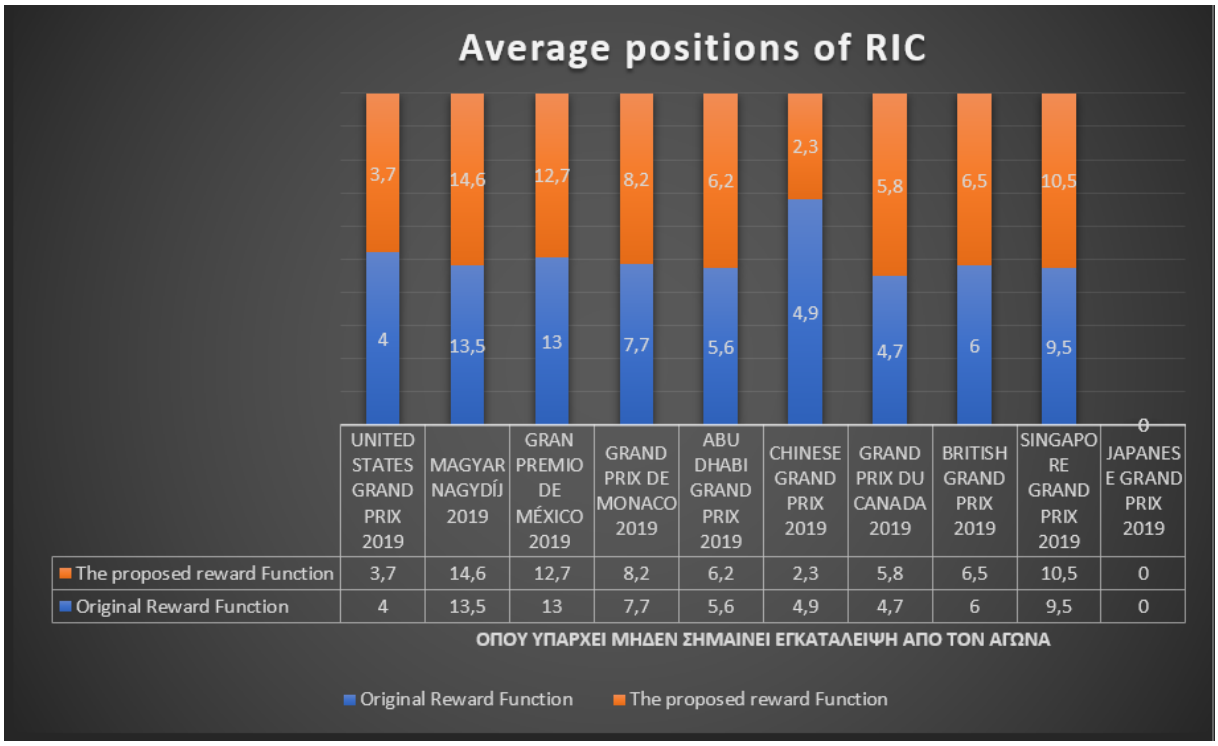
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Räikkönen είναι ένας καλός οδηγός και είχε στη διάθεσή του ένα ικανό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.17 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Räikkönen μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.17 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.17: Μέσοι όροι Kimi Räikkönen

5.1.15 Αποτελέσματα προσομοιώσεων Daniel Ricciardo

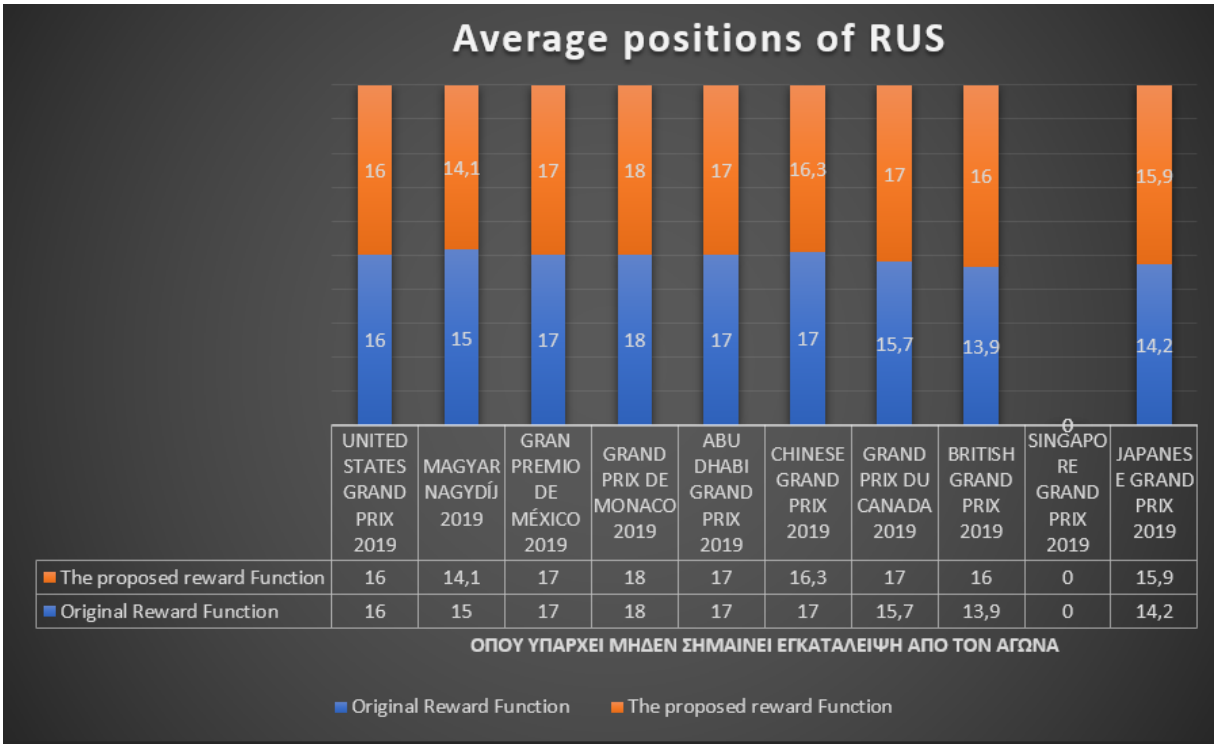
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Ricciardo είναι ένας πολύ καλός οδηγός και είχε στη διάθεσή του ένα κακό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.18 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Ricciardo μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.18 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.18: Μέσοι όροι Daniel Ricciardo

5.1.16 Αποτελέσματα προσομοιώσεων George Russell

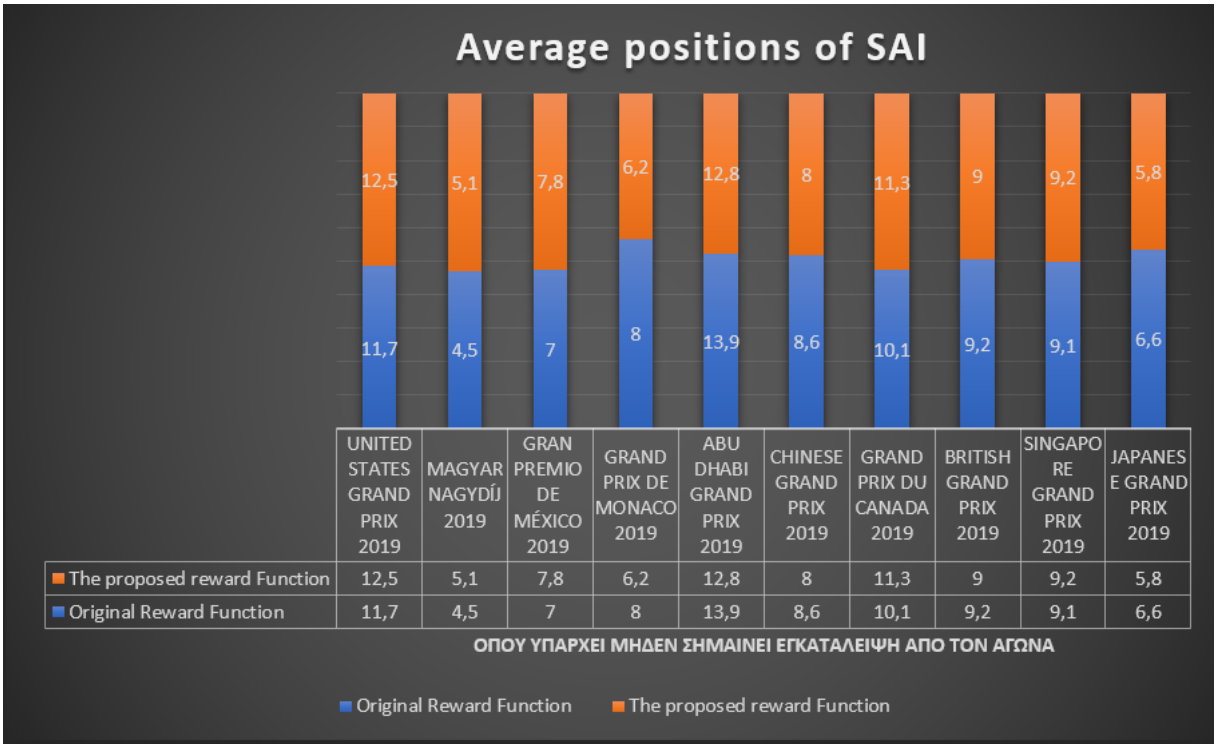
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Russell είναι ένας πολύ καλός οδηγός και είχε στη διάθεσή του ένα πολύ κακό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.19 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Russell μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.19 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.19: Μέσοι όροι George Russell

5.1.17 Αποτελέσματα προσομοιώσεων Carlos Sainz

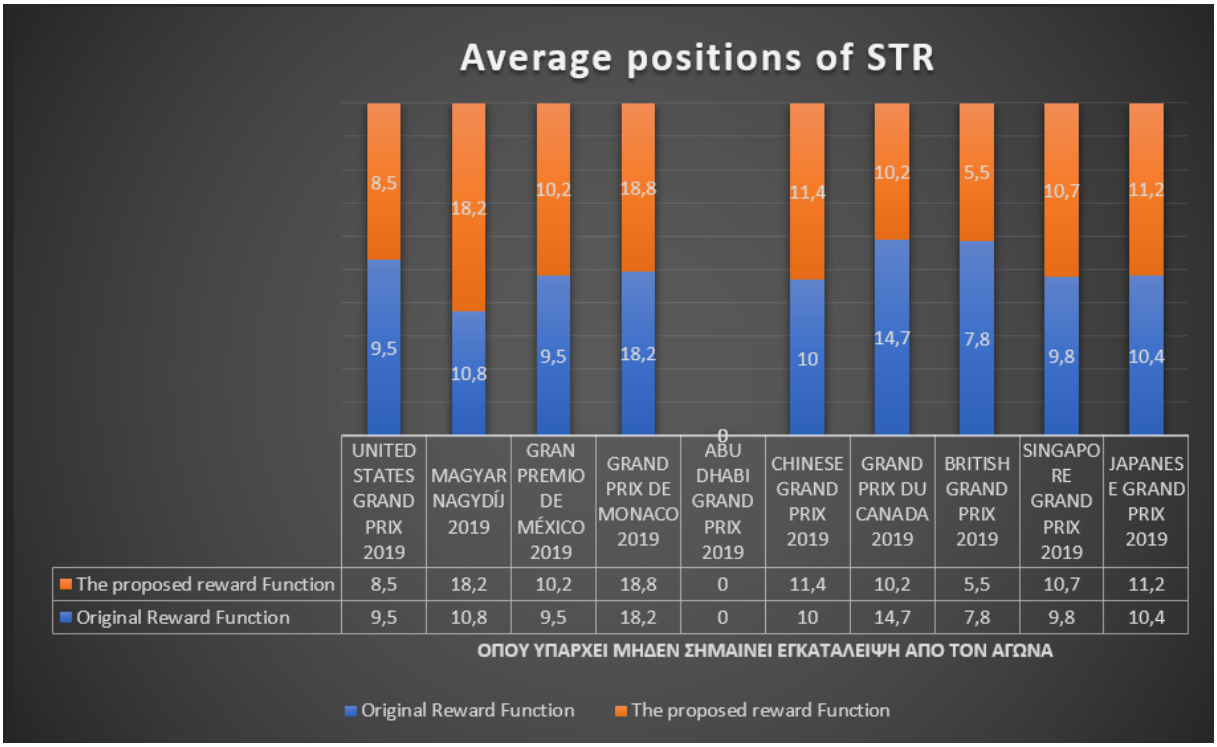
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Sainz είναι ένας καλός οδηγός και είχε στη διάθεσή του ένα ανταγωνηστικό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.20 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Sainz μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.20 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.20: Μέσοι όροι Carlos Sainz

5.1.18 Αποτελέσματα προσομοιώσεων Lance Stroll

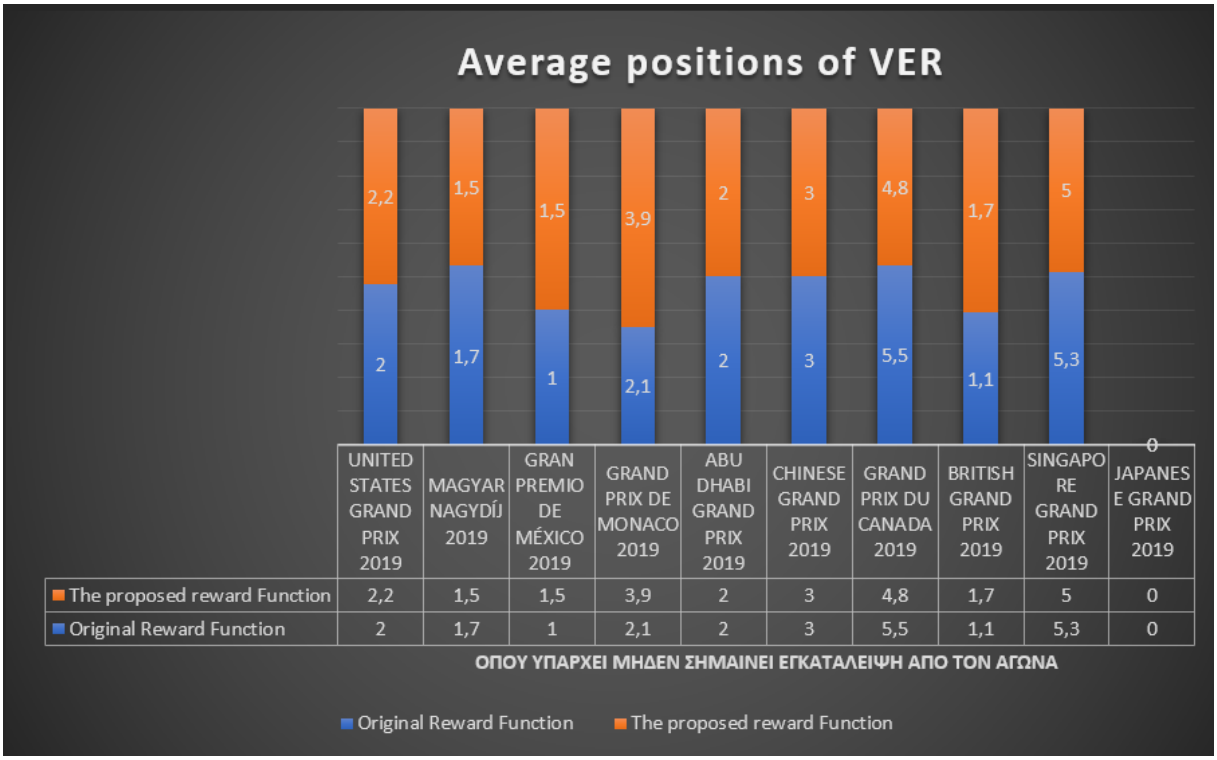
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Stroll είναι ένας μέτριος οδηγός και είχε στη διάθεσή του ένα καλό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.21 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Stroll μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.21 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.21: Μέσοι όροι Lance Stroll

5.1.19 Αποτελέσματα προσομοιώσεων Max Verstappen

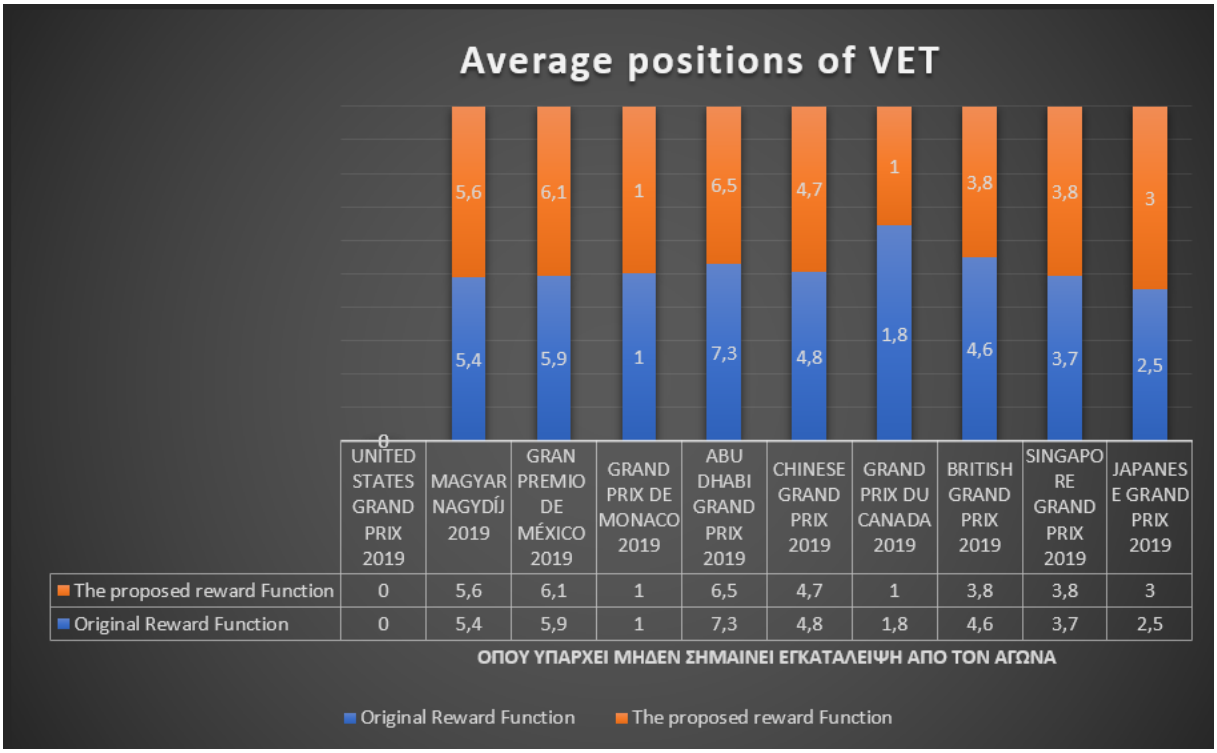
Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Verstappen είναι ένας πολύ καλός οδηγός και είχε στη διάθεσή του ένα καλό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.22 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Verstappen μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.22 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.22: Μέσοι όροι Max Verstappen

5.1.20 Αποτελέσματα προσομοιώσεων Sebastian Vettel

Με βάση τα δεδομένα που έχουμε στη διάθεσή μας, ο οδηγός Vettel είναι ένας πολύ καλός οδηγός και είχε στη διάθεσή του ένα καλό μονοθέσιο την περίοδο 2019 της Formula 1. Στο Σχήμα 5.23 μπορούμε να δούμε τον μέσο όρο θέσης που είχε ο Vettel μετά τις δέκα προσομοιώσεις σε κάθε μία από τις δέκα πίστες της περιόδου 2019 της Formula 1. Το πορτοκαλί κομμάτι του Σχήματος 5.23 αναφέρεται στον μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από την προτεινόμενη συνάρτηση reward για τον κάθε αγώνα και το μπλε κομμάτι αναφέρεται στο μέσο όρο θέσης του οδηγού έπειτα από την προσομοίωση του μοντέλου που δημιουργήθηκε από τη συνάρτηση των δημιουργών του προγράμματος.



Σχήμα 5.23: Μέσοι όροι Sebastian Vettel

5.2 Συμπεράσματα

Αφού έγινε η εκπαίδευση των μοντέλων και με τις δύο συναρτήσεις reward, εκμεταλλευτήκαμε τα μοντέλα στο δεύτερο μέρος του προγράμματος. Από τα αποτελέσματα προέκυψε ότι δεν υπήρχε μία συνάρτηση η οποία έθετε σταθερά την καλύτερη δυνατή θέση για τον κάθε οδηγό για τον οποίο γινόταν η προσομοίωση. Υπήρχαν προσομοιώσεις αγώνων όπου τα μοντέλα που προκύπταν από τη συνάρτηση των δημιουργών κατέληγαν σε καλύτερο αποτέλεσμα κατά μήκος των προσομοιώσεων, ενώ σε άλλες περιπτώσεις τα μοντέλα που είχαν προκύψει από την προτεινόμενη συνάρτηση reward είχαν καλύτερο αποτέλεσμα. Κανένα από τα δύο μοντέλα δεν κατάφερε να αντιστρέψει το αποτέλεσμα εγκατάλειψης ενός οδηγού, καθώς η εγκατάλειψη του καθενός βασίζεται σε πραγματικά δεδομένα και δεν μπορεί να αντιστραφεί.

Αν θελήσουμε ωστόσο να ξεχωρίσουμε κάποια συνάρτηση ως καλύτερη όσον αφορά τους μέσους όρους που προκύπτουν από τις προσομοιώσεις, αυτή είναι η συνάρτηση των δημιουργών. Πιο συγκεκριμένα η συνάρτηση ήταν καλύτερη στο να κατατάξει 9 από τους 20 οδηγούς στην καλύτερη δυνατή θέση. Είχε σταθερά καλύτερο μέσο όρο θέσεων σε τουλάχιστον 5 από τους αγώνες τους οποίους εξήγαμε

μέσω των προσομοιώσεων. Αν λάβουμε υπόψιν ότι μερικοί οδηγοί είχαν εγκαταλείψεις σε κάποιους από τους αγώνες τότε μπορούμε να συμπεράνουμε ότι συνάρτηση των δημιουργών είχε μία καλή επίδοση απέναντι στην προτεινόμενη συνάρτηση. Η προτεινόμενη συνάρτηση κατάφερε να είναι καλύτερη σε 3 από τους 20 οδηγούς. Αυτό το αποτέλεσμα ωστόσο προέκυψε λόγο ότι περιλαμβάνονται περισσότεροι έλεγχοι οι οποίοι σε έναν πραγματικό αγώνα παίζουν καθοριστικό ρόλο για την επιλογή στρατηγικής από τους μηχανικούς και το οποίο θα επηρεάσει άμεσα και την τελική κατάταξη του κάθε οδηγού. Όλοι αυτοί οι έλεγχοι έχουν επιπτώσεις στην τελική κατάταξη του οδηγού καθώς αν τιμωρείται συνεχώς εντός συνάρτησης reward αυτό έχει και ως αποτέλεσμα στην κατάταξη του οδηγού σε μία λιγότερο καλή θέση από αυτή που θα κατέληγε με τη συνάρτηση του δημιουργού. Επόμενο αποτέλεσμα που μπορούμε να εξάγουμε είναι ότι σε 8 από τους 20 οδηγούς οι δύο συναρτήσεις ήταν ισόπαλες σε αποδοτικότητα. Αυτό προκύπτει καθώς είχαμε ισόπαλες φορές στις οποίες ο μέσος όρος μίας συνάρτησης ήταν καλύτερος από την άλλη. Δηλαδή στους 10 αγώνες στους οποίες έγιναν οι προσομοιώσεις, 5 φορές είχε καλύτερο αποτέλεσμα το μοντέλο που δημιουργήθηκε από την προτεινόμενη συνάρτηση και 5 φορές είχε το μοντέλο της συνάρτησης των δημιουργών. Η ισοπαλία των συναρτήσεων παρατηρήθηκε σε αγώνες στους οποίους παίζει σημαντικό ρόλο η δυναμική του μονοθέσιου όπως για παράδειγμα η πίστα στις Ηνωμένες Πολιτείες της Αμερικής. Στις προσομοιώσεις μας οι 6 από αυτές τις πίστες βασίζονται αρκετά στη δυναμική του μονοθέσιου. Η ισοπαλία συνέβη καθώς σε αυτές τις πίστες οι στρατηγικές που επιλέγουν τα δύο μοντέλα παίζουν μικρότερο ρόλο από ότι στις υπόλοιπες. Εκτός από τη στρατηγική που επιλέγουν τα μοντέλα εδώ παίζει σημαντικό ρόλο και η δυναμική του μονοθέσιου, δηλαδή πόσο πιο γρήγορο είναι από τα υπόλοιπα και έτσι η προσπάθεια για την κατάταξη του οδηγού στην καλύτερη δυνατή θέση εξαρτάται αρκετά από αυτό, χωρίς όμως η στρατηγική που επιλέγεται να μην είναι εξίσου σημαντική, απλώς όχι στον ίδιο βαθμό σε σχέση με τις υπόλοιπες πίστες όπου ένα λάθος στη στρατηγική μπορεί να βάλει τον οδηγό αρκετές θέσεις πιο κάτω από ότι ήταν πριν το pit-stop του.

Κεφάλαιο 6

Συμπεράσματα και Μελλοντικές Προεκτάσεις

Η δυναμική που είχε το συγκεκριμένο πρόγραμμα στο οποίο εργαστήκαμε φάνηκε από την αρχή καθώς λαμβάνονται υπόψιν αρκετά και λεπτομερή δεδομένα, όπως και εξωτερικοί παράγοντες ενός αγώνα Formula 1. Η συγκεκριμένη μορφή του προγράμματος του δίνει τη δυνατότητα να εφαρμοστεί και σε άλλες κατηγορίες του μηχανοκίνητου αθλητισμού και ακόμα να εξελιχθεί ώστε να λαμβάνει υπόψιν περισσότερες παραμέτρους που απαιτεί κάθε κατηγορία.

Η εφαρμογή τέτοιων προγραμμάτων μπορεί να βοηθήσει μηχανικούς αλλά και όλη την ομάδα στο να εκμεταλλευτούν μεγάλο όγκο δεδομένων. Μέσω της μηχανικής μάθησης να βρουν σημεία του αυτοκινήτου τους τα οποία χρειάζονται εξέλιξη, να γίνουν πιο παραγωγικοί εστιάζοντας μόνο σε αυτά τα κομμάτια που χρειάζονται βελτίωση και να έχουν τη δυνατότητα εντός του αγωνιστικού χώρου να εξερευνήσουν δεδομένα αντίπαλων ομάδων και να τα συγκρίνουν με τα δικά τους. Με αυτό τον τρόπο τοποθετούνται σε πλεονεκτική θέση ώστε να εκμεταλλευτούν τις αδυναμίες των αντιπάλων τους και να εξελίξουν τις δικές. Έτσι κάθε ομάδα που θα χρησιμοποιεί μηχανική θα είναι πάντα ένα βήμα μπροστά, θα λαμβάνει πληροφορίες όταν και όποτε της είναι αναγκαίες, γλιτώνοντάς τους χρόνο από περιττές έρευνες. Έτσι μπορούν να εστιάσουν περισσότερο στην κατάκτηση της πρώτης θέσης που είναι και ο στόχος των περισσότερων ομάδων.

Η μηχανική μάθηση θα παίξει βασικό ρόλο στην εξέλιξη των οχημάτων που έχει η κάθε ομάδα, αφού στις μέρες μας οι περισσότερες διοργανώσεις έχουν ως κύριο στόχο τους, τη μείωση των χρημάτων που δαπανά μία ομάδα για δοκιμές εξέλιξης

το οποίο έχει ως αποτέλεσμα να μειώνεται ο χρόνος που μπορούν να περνούν στην πίστα οι ομάδες για να εξελίσσουν το όχημά τους. Επομένως θα τους δοθεί η ευκαιρία να είναι αρκετά προετοιμασμένοι πριν να αγωνιστούν στον πραγματικό αγώνα.

Επιπλέον, η μηχανική μάθηση θα τους βοηθήσει στην επιλογή του καταλληλότερου οδηγού, αφού αναλύοντας τα δεδομένα και τα προηγούμενα κατορθώματά του θα μπορέσουν να στραφούν στην καταλληλότερη επιλογή χωρίς να χρειάζεται να παρακολουθούν τους αγώνες στους οποίους συμμετέχει. Τέλος, η μηχανική μάθηση θα μπορέσει να συμβάλει στην καλύτερη εμπειρία των θεατών, αφού θα είναι ικανή να κατανοήσει ακριβώς το περιεχόμενο που ανταποκρίνεται στις αρεσκείες των οπαδών οπότε οι διοργανώσεις θα έχουν τη δυνατότητα να βελτιώσουν την εμπειρία του θεατή, ο οποίος είναι αναπόσπαστο κομμάτι κάθε αγώνα που διοργανώνεται.

Ήδη υπάρχουν αρκετές παραλλαγές οι οποίες εφαρμόζονται σε πραγματικούς αγώνες του μηχανοκίνητου αθλητισμού ή μένει να εφαρμοστούν και να δώσουν για πάντα μία διαφορετική προσέγγιση στη φιλοσοφία και οργάνωση ομάδας και οχήματος σε κάθε διαφορετική κατηγορία του μηχανοκίνητου αθλητισμού.

Βιβλιογραφία

[1] I. El Naqa and M. J. Murphy, "What is machine learning?," in *Machine Learning in Radiation Oncology*, pp. 3–11, Springer, 2015.

[2] I. C. Education, "What is supervised learning?." https://www.ibm.com/cloud/learn/supervised-learningl, 2020.

[3] L. H. . M. N. Office, "Explained: Neural networks." https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414, 2017.

[4] . Ματθαίου, "Machine learning." https://nowmag.gr/machine-learning/, 2020.

[5] M. A. Wiering and M. Van Otterlo, "Reinforcement learning," *Adaptation, Learning, and Optimization*, vol. 12, no. 3, p. 729, 2012.

[6] P. Verma and S. Diamantidis, "What is reinforcement learning?." https://www.synopsys.com/ai/what-is-reinforcement-learning.html, 2021.

[7] M. Wang, "Deep q-learning tutorial: mindqn." https://towardsdatascience.com/deep-q-learning-tutorial-mindqn-2a4c855abffc, 2020.

[8] S. Mondal, "Q-network reinforcement learning model." https://medium.com/analytics-vidhya/q-network-reinforcement-learning-model-fe3f8d982aec, 2020.

[9] A. Choudhary, "A hands-on introduction to deep q-learning using openai gym in python." https://www.analyticsvidhya.com/blog/2019/04/introduction-deep-q-learning-python/, 2020.

[10] DEEPLIZARD, "Reinforcement learning - developing intelligent agents." https://deeplizard.com/learn/video/Bcuj2fTH4_4, 2018.

[11] R. S. S. Shangtong Zhang, "Reinforcement learning - developing intelligent agents." https://arxiv.org/pdf/1712.01275.pdf, 2018.

[12] A. B. RS Sutton, "Reinforcement learning: An introduction." https://d1wqtxts1xzle7.cloudfront.net/38529120/9780262257053_index-with-cover-page-v2.pdf?Expires=1664381653&Signature=Mm4Yza5HTpR-7mpui8OiyaeQHoHdcKUYPEUOG3Z7~ViHsaYtt5QwuU3hZdEUluPomRn~ms1EyxYl5lXWL5UNqj9RdLEfR_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA, 2018.

[13] MLK, "Rbeginner's guide to policy in reinforcement learning." https://machinelearningknowledge.ai/beginners-guide-to-what-is-policy-in-reinforcement-learning/, 2021.

[14] R. SAGAR, "On-policy vs off-policy reinforcement learning." https://analyticsindiamag.com/reinforcement-learning-policy/, 2020.

[15] H. Hristov, "Off-policy vs. on-policy reinforcement learning." https://www.baeldung.com/cs/off-policy-vs-on-policy, 2022.

[16] D. Precup, "Exploration and exploitation in reinforcement learning." https://neuro.bstu.by/ai/To-dom/My_research/Papers-2.1-done/RL/0/FinalReport.pdf, 2004.

[17] M. Piñol, "Exploration versus exploitation dilemma in reinforcement learning." https://www.cs.csustan.edu/~mmartin/teaching/CS4960S15/Pinol-Senior%20Seminar.pdf, 2015.

[18] A. Ajagekar, "Adam." https://optimization.cbe.cornell.edu/index.php?title=Adam, 2021.

[19] D. P. K. J. L. Ba, "Adam: A method for stochastic optimization." https://arxiv.org/pdf/1412.6980.pdf, 2015.

[20] S. Doshi, "Various optimization algorithms for training neural network." https://towardsdatascience.com/optimizers-for-training-neural-network-59450d71caf6, 2019.

[21] G. Neto, "From single-agent to multi-agent reinforcement learning: Foundational concepts and methods," *Learning Theory Course*, vol. 2, 2005.

[22] Y. Zhang, Q. Yang, D. An, and C. Zhang, "Coordination between individual agents in multi-agent reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 11387–11394, 2021.

[23] F. James, "Monte carlo theory and practice," *Reports on Progress in Physics*, vol. 43, no. 9, p. 1145, 1980.

[24] S. Raychaudhuri, "Introduction to monte carlo simulation," in *2008 Winter Simulation Conference*, pp. 91–100, IEEE, 2008.

[25] J. Brownlee, "How to configure the learning rate when training deep learning neural networks." hhttps://machinelearningmastery.com/learning-rate-for-deep-learning-neural-networks/, 2019.

[26] X. Lin, Q. Xing, and F. Liu, "Choice of discount rate in reinforcement learning with long-delay rewards," *Journal of Systems Engineering and Electronics*, vol. 33, no. 2, pp. 381–392, 2022.

[27] P. H. Moreno, "Optimize f1 aerodynamic geometries via design of experiments and machine learning." https://aws.amazon.com/blogs/machine-learning/optimize-f1-aerodynamic-geometries-via-design-of-experiments-and-machine-learning/, 2022.

[28] S. A. Löckel, "Machine learning for modeling and analyzing of race car drivers." https://tuprints.ulb.tu-darmstadt.de/20218/, 2022.

[29] M. Jaritz, R. De Charette, M. Toromanoff, E. Perot, and F. Nashashibi, "End-to-end race driving with deep reinforcement learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2070–2075, IEEE, 2018.

[30] G. L. Gil Gómez, M. Nybacka, L. Drugge, and E. Bakker, "Machine learning to classify and predict objective and subjective assessments of vehicle dynamics: the case of steering feel," *Vehicle System Dynamics*, vol. 56, no. 1, pp. 150–171, 2018.

[31] C. S. Pedroso, "Using machine learning to predict if a f1 driver will score." https://towardsdatascience.com/machine-learning-to-predict-if-a-f1-driver-will-score-e741f057494d, 2021.

[32] Acronis, "Toyota gazoo racing picks to deliver artificial intelligence and machine learning." https://motorsport.tech/technology/toyota-gazoo-racing-benefitting-from-acronis-to-deliver-ai-and-machine-learning, 2021.

[33] P. C. X. M. L. T. Y. Sun1 and M. Liu, "High-speed autonomous drifting with deep reinforcement learning." https://arxiv.org/pdf/2001.01377.pdf, 2020.

[34] J. Gooch, "Ai takes its place in the race." https://www.ansys.com/blog/ai-takes-its-place-in-the-race, 2021.

[35] T. U. of Munich, "Tum's ai-controlled race car wins the indy autonomous challenge." https://www.tum.de/en/news-and-events/all-news/press-releases/details?no_cache=1&tx_news_pi1%5baction%5d=detail&tx_news_pi1%5bcontroller%5d=News&tx_news_pi1%5bnews%5d=37010, 2021.

[36] AWS, "Why f1 chooses aws." https://aws.amazon.com/f1/, 2022.

[37] A. Heilmeier, M. Graf, and M. Lienkamp, "A race simulation for strategy decisions in circuit motorsports," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2986–2993, 2018.