



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών

και Μηχανικών Υπολογιστών

Τομέας Επικοινωνιών, Ηλεκτρονικής Και
Συστημάτων Πληροφορικής

Μηχανική Μάθηση για Έξυπνα Δίκτυα Ενέργειας-Vehicle-to-Grid-Services

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΠΑΝΑΓΙΩΤΗΣ Ν ΛΟΥΛΑΚΗΣ

Επιβλέπων: Εμμανουήλ Βαρβαρίγος

Καθηγητής Ε.Μ.Π

Αθήνα, Οκτώβριος 2021



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών

και Μηχανικών Υπολογιστών

Τομέας Επικοινωνιών, Ηλεκτρονικής Και
Συστημάτων Πληροφορικής

Μηχανική Μάθηση για Έξυπνα Δίκτυα Ενέργειας-Vehicle-to-Grid-Services

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΠΑΝΑΓΙΩΤΗΣ Ν ΛΟΥΛΑΚΗΣ

Επιβλέπων: Εμμανουήλ Βαρβαρίγος

Καθηγητής Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 18^η Οκτωβρίου 2021

Αθήνα, Οκτώβριος 2021

.....

Εμμανουήλ Βαρβαρίγος

Καθηγητής Ε.Μ.Π

.....

Θεοδώρα Βαρβαρίγου

Καθηγήτρια Ε.Μ.Π

.....

Συμεών Παπαβασιλείου

Καθηγητής Ε.Μ.Π

.....

Παναγιώτης Ν Λουλάκης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π

Copyright © Παναγιώτης Λουλάκης, 2021 Με επιφύλαξη παντός δικαιώματος. All rights reserved. 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 δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Σκοπός αυτής της εργασίας είναι η μελέτη και σύγκριση μεθόδων τεχνητής νοημοσύνης σε δίκτυα ελεγχόμενης και αμφίδρομης φόρτισης ηλεκτρικών αυτοκινήτων. Συγκεκριμένα, θεωρήσαμε ένα μοντέλο σταθμού φόρτισης και parking ηλεκτρικών αυτοκινήτων το οποίο εφαρμόζει τεχνικές vehicle-to-grid. Στη συνέχεια δοκιμάσαμε τη χρήση διαφορετικών αλγορίθμων ενισχυτικής μάθησης από όλες τις κατηγορίες (διακριτούς, συνεχείς και πολυδιάστατους) τους οποίους συγκρίναμε τόσο με το χρηματικό κέρδος που αποδίδουν όσο και στην φθορά της μπαταρίας των αυτοκινήτων που προκαλούν. Η μελέτη αυτή έγινε σε μεταβλητά μεγέθη parking καθώς και σε διαφορετικές κατανομές άφιξης αυτοκινήτων. Τέλος όλα τα αποτελέσματα συγκρίθηκαν με μια παραδοσιακή πολιτική φόρτισης, η οποία δεν κάνει χρήση μεθόδων τεχνητής νοημοσύνης, και αποδείξαμε ότι η χρήση κατάλληλων αλγορίθμων μηχανικής μάθησης μπορεί να φέρει πολύ καλύτερα αποτελέσματα.

Λέξεις-κλειδιά: Κλιματική αλλαγή, Ανανεώσιμες πηγές ενέργειας, Ηλεκτρικά αυτοκίνητα, Φθορά μπαταρίας, Συστήματα ελεγχόμενης φόρτισης, Ενισχυτική μάθηση, Βαθιά Q-μάθηση, Ανεξάρτητη Q-μάθηση, Ελαφρύς κριτής ηθοποιός, Μαρκοβιανή διαδικασία απόφασης

Abstract

The goal of this thesis is to study and compare different machine learning techniques when applied to controlled, two-way electric car charging systems. Firstly, we introduced an electric car parking and charging station which applied vehicle-to-grid techniques. Then we tested different reinforcement learning algorithms from multiple categories (discrete, continuous, multi-dimensional) and compared them based on both the monetary profit they managed to acquire, and the battery degradation they caused. This study was then repeated on different parking station sizes and with different vehicle arrival distributions. Finally, the algorithms were compared with a traditional charging policy, which doesn't utilize artificial intelligence, in order to prove the advantages of utilizing machine learning algorithms to achieve better results.

Key-words: Climate change, Renewable energy sources, Battery Degradation, Controlled charging systems, Reinforcement learning, Deep Q network, Independent Q learning, Soft actor critic, Markovian Decision process.

Ευχαριστίες

Στο σημείο αυτό, καθώς αυτή η εργασία σηματοδοτεί το τέλος των προπτυχιακών μου σπουδών, θα ήθελα να ευχαριστήσω ορισμένους ανθρώπους για την βοήθεια που μου πρόσφεραν καθ' όλη τη διάρκεια των τελευταίων χρόνων.

Αρχικά θέλω να ευχαριστήσω τους καθηγητές μου για τις γνώσεις που μου παρείχαν τα τελευταία 5 χρόνια, και κυρίως τον κ. Εμμανουήλ Βαρβαρίγο ο οποίος, ως επιβλέπων της διπλωματικής εργασίας, μου έδωσε την ευκαιρία να ασχοληθώ με αυτό το αντικείμενο. Επιπλέον θα ήθελα να ευχαριστήσω τον διδάκτορα Γεώργιο Τσαούσογλου και την υποψήφια διδάκτορα Αικατερίνη Μητροπούλου για τη βοήθεια που μου πρόσφεραν τόσο στην κατανόηση του θέματος όσο και στη πραγματοποίησή του.

Επιπρόσθετα, θα ήθελα να ευχαριστήσω τους φίλους μου, Αντώνιο Βαλμά, Ιωσήφ-Γεώργιο Παναγιωτόπουλο και Ισαβέλλα Κουκουλά για την συμπαράστασή τους και τις ατέλειωτες στιγμές γέλιου που μου παρείχαν αυτά τα 5 χρόνια.

Τέλος θα ήθελα να ευχαριστήσω την οικογένεια μου για την υποστήριξη που μου χάρισαν σε όλη τη ζωή μου. Συγκεκριμένα θέλω να ευχαριστήσω τον αδελφό μου Ιωάννη Άγγελο Λουλάκη, τον πατέρα μου Νικόλαο Λουλάκη και την μητέρα μου Ευγενία Παπαθανασίου.

Περιεχόμενα

Περίληψη.....	5
Abstract	7
Ευχαριστίες.....	9
Κατάλογος Εικόνων	13
Κατάλογος Πινάκων	14
Συνομογραφίες	15
Εισαγωγή	16
Αντικείμενο της εργασίας	17
Δομή της Εργασίας	19
Κεφάλαιο 1, Κλιματική αλλαγή	23
Ορισμός Κλιματικής Αλλαγής.....	24
Επιπτώσεις της Κλιματικής Αλλαγής στην Οικονομία και την Υγεία	24
Συνεισφορά της Ενέργειας στην Κλιματική αλλαγή	25
Κεφάλαιο 2, Ανανεώσιμες Πηγές Ενέργειας.....	28
Ορισμός	29
Πλεονεκτήματα και μειονεκτήματα.....	29
Χρήση των Ανανεώσιμων Πηγών Ενέργειας για την Αντιμετώπιση της Κλιματικής Αλλαγής	30
Κεφάλαιο 3, Ηλεκτρικά αυτοκίνητα.....	33
Ηλεκτρικά Αυτοκίνητα και Κλιματική Αλλαγή	34
Είδη Μπαταριών Ηλεκτρικών Αυτοκινήτων.....	34
Φθορά Μπαταριών Ηλεκτρικών Αυτοκινήτων	36
Κεφάλαιο 4, Αγορά Ηλεκτρικής Ενέργειας	41
Το Ηλεκτρικό Δίκτυο.....	42
Το Κόστος Ηλεκτρικής Ενέργειας	43
Κεφάλαιο 5, Vehicle-to-Grid	46
Ορισμός	47
Πλεονεκτήματα του Vehicle-to-Grid	48
Μειονεκτήματα και Προκλήσεις του Vehicle-to-Grid.....	50
Vehicle-to-Grid και Τεχνητή Νοημοσύνη	51
Κεφάλαιο 6, Ενισχυτική Μάθηση.....	54
Ορισμός	55
Μαθηματικό υπόβαθρο	57
Αλγόριθμοι Ενισχυτικής Μάθησης.....	60

DQN	60
IQL.....	64
SAC.....	67
Κεφάλαιο 7, Φορμαλισμός του Προβλήματος.....	74
Ορισμός του προβλήματος	75
Σχεδίαση Μοντέλου	78
Παραγωγή δεδομένων	81
Αλγόριθμοι Απόφασης.....	83
Βασικός αλγόριθμος.....	83
IQL.....	85
SAC.....	86
Κεφάλαιο 8, Πειράματα και Αποτελέσματα	88
Παράθεση των πειραμάτων.....	89
Πείραμα Ίσης κατανομής	90
Πείραμα κατανομής Περισσότερων Αυτοκινήτων.....	91
Πείραμα κατανομής Λιγότερων Αυτοκινήτων	92
Τελικά αποτελέσματα	94
Βιβλιογραφία	96

Κατάλογος Εικόνων

Εικόνα 1: Χρονική μεταβολή της Κατάσταση Υγείας της μπαταρίας ενός ηλεκτρικού αυτοκινήτου.....	36
Εικόνα 2: Μείωση μέγιστης χωρητικότητας μπαταρίας ηλεκτρικών αυτοκινήτων, σε διαφορετικά επίπεδα φόρτισης και θερμοκρασίας, σε διάστημα 900 ημερών	38
Εικόνα 3: Τιμές κόστους ηλεκτρικού ρεύματος ανά ώρα	43
Εικόνα 4: Μοντέλο Ενισχυτικής μάθησης	55
Εικόνα 5: Q-learning algorithm [68]	61
Εικόνα 6: DQN algorithm[72].....	63
Εικόνα 7: Reinforce Algorithm.....	68
Εικόνα 8: Q Actor Critic Algorithm.....	69
Εικόνα 9: Soft Actor Critic Algorithm[79]	72
Εικόνα 10: Κόστος ηλεκτρικού ρεύματος ανά χρονική στιγμή	81
Εικόνα 11: Κατανομή αφίξεων ηλεκτρικών αυτοκινήτων.....	82
Εικόνα 12: Αφίξεις ηλεκτρικών αυτοκινήτων ανά χρονική στιγμή	82
Εικόνα 13: Επιβράβευση λόγω αγοράς ηλεκτρικού ρεύματος σε πείραμα ίσης κατανομής.....	90
Εικόνα 14: Επιβράβευση λόγω φθοράς της μπαταρίας σε πείραμα ίσης κατανομής.....	90
Εικόνα 15: Συνολική επιβράβευση σε πείραμα ίσης κατανομής.....	90
Εικόνα 16: Επιβράβευση λόγω αγοράς ηλεκτρικού ρεύματος σε πείραμα μεγαλύτερης κατανομής.....	91
Εικόνα 17: Επιβράβευση λόγω φθοράς της μπαταρίας σε πείραμα ίσης κατανομής.....	91
Εικόνα 18: Συνολική επιβράβευση σε πείραμα μεγαλύτερης κατανομής	92
Εικόνα 20: Επιβράβευση λόγω φθοράς της μπαταρίας σε πείραμα μικρότερης κατανομής.....	93
Εικόνα 19: Επιβράβευση λόγω αγοράς ηλεκτρικού ρεύματος σε πείραμα μικρότερης κατανομής.....	93
Εικόνα 21: Συνολική επιβράβευση σε πείραμα μικρότερης κατανομής	93

Κατάλογος Πινάκων

Πίνακας 1: Τάση ηλεκτρικής μπαταρίας σε διαφορετικά επίπεδα φόρτισης	38
Πίνακας 2: απαραίτητοι κύκλοι φόρτισης μέχρι το τέλος ζωής μιας ηλεκτρικής μπαταρίας για διαφορετικούς ρυθμούς φόρτισης	78
Πίνακας 3: Απώλεια χωρητικότητας ηλεκτρικής μπαταρίας σε διαφορετικά επίπεδα φόρτισης.....	78
Πίνακας 4: Πλήθος νευρώνων κάθε επιπέδου του νευρωνικού δικτύου του αλγορίθμου DQN	84
Πίνακας 5: Πλήθος νευρώνων κάθε επιπέδου του νευρωνικού δικτύου του αλγορίθμου IQL	85

Συντομογραφίες

- ΑΠΕ: Ανανεώσιμες πηγές Ενέργειας
- V2G: Vehicle-to-Grid
- ΑΕΠ: Ακαθάριστο Εθνικό Προϊόν
- ΠΟΥ: Παγκόσμιος Οργανισμός υγείας
- SOH: State Of Health
- SOC: State Of Charge
- MDP: Markov Decision Process
- DQN: Deep Q-Network
- IQL: Independent Q-Learning
- SAC: Soft Actor Critic

Εισαγωγή

Αντικείμενο της εργασίας

Ένα από τα μεγαλύτερα προβλήματα που αντιμετωπίζει η επιστήμη τον 21^ο αιώνα είναι το φαινόμενο της κλιματικής αλλαγής. Πέρα από τις επιπτώσεις που αυτή επιφέρει στο περιβάλλον [1] υπάρχουν πολύ σοβαρές συνέπειες τόσο στην υγεία[2] όσο και στην οικονομία[3]. Ένας από τους σημαντικότερους παράγοντες που επιτείνει την κλιματική αλλαγή είναι η βιομηχανία της ενέργειας και πιο συγκεκριμένα η εκτενής χρήση μη ανανεώσιμων πηγών ενέργειας [4]. Για την αντιμετώπιση αυτού του φαινομένου υπογράφηκε η Ευρωπαϊκή Πράσινη Συμφωνία[5] η οποία θέτει ως στόχο την αύξηση της χρήσης των ανανεώσιμων πηγών ενέργειας[6].

Οι ανανεώσιμες πηγές ενέργειας όμως έχουν και αυτές κάποια σοβαρά μειονεκτήματα. Ένα από αυτά είναι η μη ελεγχόμενη παραγωγή[7]. Επειδή η παραγωγή των ανανεώσιμων πηγών ενέργειας εξαρτάται από το περιβάλλον, δεν μπορεί να ρυθμιστεί από το δίκτυο τόσο εύκολα όσο αυτή των μη ανανεώσιμων πηγών. Για το λόγο αυτό υπάρχει πολύ μεγάλη ανάγκη αποθήκευσης της ηλεκτρικής ενέργειας - όταν η παραγωγή της είναι περισσότερη από την αντίστοιχη κατανάλωση- ώστε να χρησιμοποιείται στην αντίθετη περίπτωση. Αυτή η ιδιαιτερότητα έχει ως αποτέλεσμα και την μεγαλύτερη διασπορά του κόστους της ενέργειας ανά ώρα[8]. Μια μέθοδος που μπορεί να εφαρμοστεί τόσο για την κάλυψη της ανάγκης ηλεκτρικών μπαταριών όσο και για την οικονομικότερη χρήση ενέργειας μεταβλητού κόστους είναι η μέθοδος Vehicle-to-Grid(V2G).

Ο όρος Vehicle-to-Grid περιγράφει ένα σύστημα ηλεκτρικών αυτοκινήτων στο οποίο οι μπαταρίες τους χρησιμοποιούνται από το ηλεκτρικό δίκτυο, δηλαδή το αυτοκίνητο δίνει ρεύμα στο δίκτυο από την μπαταρία του. Με την χρήση αυτής της μεθόδου ένα parking ηλεκτρικών

αυτοκινήτων μπορεί να δρα ως συσσωρευτής (aggregator) και να προσαρμόζει την φόρτιση των αυτοκινήτων ώστε να παρέχει ενέργεια στο δίκτυο -όταν χρειάζεται- και να φορτίζει τα αυτοκίνητα με μικρότερο κόστος. Ένας σημαντικός περιορισμός στις αποφάσεις του συσσωρευτή φυσικά είναι οι ανάγκες του ιδιοκτήτη του κάθε αυτοκινήτου, δηλαδή όταν αυτός επιστρέψει στο αυτοκίνητο του, αυτό να έχει αρκετά ψηλό επίπεδο φόρτισης ώστε να ικανοποιήσει τις ανάγκες του.

Ένα ακόμα σημαντικό εμπόδιο στην χρήση μεθόδων V2G είναι η φθορά της μπαταρίας των ηλεκτρικών αυτοκινήτων[10]. Η φόρτιση και η χρήση της μπαταρίας ενός ηλεκτρικού αυτοκινήτου προκαλεί φθορά σε αυτή και μειώνει την αποθηκευτική της ικανότητα. Επομένως οι αποφάσεις του συσσωρευτή για την φόρτιση και την εκφόρτιση των αυτοκινήτων πρέπει να παίρνουν υπόψη όχι μόνο το κόστος του ρεύματος και τις ανάγκες του δικτύου και του ιδιοκτήτη του αυτοκινήτου αλλά και τη φθορά των μπαταριών την οποία πρέπει να προσπαθούν να διατηρούν σε όσο το δυνατόν χαμηλότερα επίπεδα.

Οι αποφάσεις του συσσωρευτή για το κάθε αυτοκίνητο μπορούν να μοντελοποιηθούν ως ένα πρόβλημα βελτιστοποίησης, όπου η λύση είναι ο βέλτιστος τρόπος φόρτισης και εκφόρτισης του κάθε αυτοκινήτου με στόχο τη μεγιστοποίηση του χρηματικού κέρδους, τη μείωση της φθοράς της μπαταρίας και την ικανοποίηση των περιορισμών που προκύπτουν από τις ανάγκες του οδηγού του αυτοκινήτου. Για την λύση αυτού του προβλήματος κάναμε χρήση αλγορίθμων μηχανικής μάθησης, και πιο συγκεκριμένα ενισχυτικής μάθησης (Reinforcement Learning). Η Ενισχυτική μάθηση είναι ένας τομέας μηχανικής μάθησης όπου πράκτορες (agents) παίρνουν αποφάσεις (actions) σε ένα

περιβάλλον (environment) με σκοπό την μεγιστοποίηση κάποιας ανταμοιβής (reward)[11]. Οι πράκτορες παρακολουθούν τις αλλαγές που προκαλούν οι πράξεις τους στο περιβάλλον καθώς και την ανταμοιβή που αυτές εισπράττουν και μαθαίνουν με βάση κάποια πολιτική (policy). Σε αυτή την εργασία έγινε χρήση πολλών μεθόδων ενισχυτικής μάθησης (Agent, Policy), σε συσσωρευτές διαφόρων μεγεθών, με σκοπό την σύγκριση αναφορικά με το κέρδος που αποδίδουν τόσο σε κατανομή άφιξης και αναχώρησης αυτοκινήτων όμοια με την προβλεπόμενη (στην οποία έγινε η εκπαίδευση τους), όσο και σε άγνωστες κατανομές.

Δομή της Εργασίας

Η δομή της διπλωματικής εργασίας περιγράφεται παρακάτω:

- **Κεφάλαιο 1, Κλιματική αλλαγή:** Γίνεται μια σύντομη παρουσίαση του φαινομένου της κλιματικής αλλαγής καθώς και των επιπτώσεών της. Επιπλέον παρατίθεται η συνεισφορά της βιομηχανίας της ενέργειας στη κλιματική αλλαγή.
- **Κεφάλαιο 2, Ανανεώσιμες πηγές Ενέργειας:** Αρχικά παρουσιάζεται ο ορισμός των ανανεώσιμων πηγών ενέργειας. Έπειτα γίνεται μια σύντομη ανασκόπηση των πλεονεκτημάτων και των μειονεκτημάτων τους σε σύγκριση με τις άλλες μορφές ενέργειας. Τέλος αναφέρεται η σημασία τους για την καταπολέμηση της κλιματικής αλλαγής.
- **Κεφάλαιο 3, Ηλεκτρικά αυτοκίνητα:** Πρωτίστως αναφέρεται η σημασία των ηλεκτρικών αυτοκινήτων στην καταπολέμηση της κλιματικής αλλαγής. Ύστερα γίνεται μια σύντομη ανασκόπηση και σύγκριση των διαφορετικών τεχνολογιών των μπαταριών των

ηλεκτρικών αυτοκινήτων. Τέλος δίνεται μια ανάλυση των παραγόντων που συμβάλουν στην φθορά των ηλεκτρικών μπαταριών.

- **Κεφάλαιο 4, Αγορά ηλεκτρικής Ενέργειας:** Δίνεται μια εικόνα της δομής του ηλεκτρικού δικτύου. Επιπλέον γίνεται μια σύντομη ανάλυση του κόστους του ηλεκτρικού ρεύματος και των παραγόντων που το επηρεάζουν.
- **Κεφάλαιο 5, Vehicle-to-Grid:** Αρχικά δίνεται ο ορισμός της τεχνολογίας Vehicle-to-grid καθώς και τα κύρια πλεονεκτήματά της. Έπειτα δίνονται τα μειονεκτήματα και οι προκλήσεις που πρέπει να αντιμετωπιστούν για την εκτεταμένη χρήση της τεχνολογίας. Τέλος γίνεται μια σύντομη αναφορά του πως μπορούν πολλά από αυτά τα προβλήματα να αντιμετωπιστούν με μεθόδους τεχνητής νοημοσύνης.
- **Κεφάλαιο 6, Ενισχυτική Μάθηση:** Ξεκινώντας γίνεται μια περιγραφή των μεθόδων ενισχυτικής μάθησης στην οποία δίνεται το μαθηματικό υπόβαθρο που είναι απαραίτητο για την κατανόησή της. Ύστερα γίνεται επεξήγηση διαφόρων αλγορίθμων ενισχυτικής μάθησης από διαφορετικές κατηγορίες.
- **Κεφάλαιο 7, Φορμαλισμός του προβλήματος:** Παρουσιάζεται η μοντελοποίηση του προβλήματος ως μια μαρκοβιανή διαδικασία καθώς και η υλοποίηση των περιβαλλόντων για τους πράκτορες της ενισχυτικής μάθησης. Επιπλέον παρουσιάζονται 3 αλγόριθμοι ενισχυτικής μάθησης με σκοπό την επίλυση του προβλήματος καθώς και ένας παραδοσιακός αλγόριθμος που χρησιμοποιείται για λόγους σύγκρισης.
- **Κεφάλαιο 8, Πειράματα και Αποτελέσματα:** Κατ' αρχάς παρουσιάζονται τα πειράματα στα οποία συγκρίθηκε η

επίδοση των αλγορίθμων που ορίστηκαν στο κεφάλαιο 7 καθώς και τα αποτελέσματά τους. Τέλος γίνεται μια συνολική ανάλυση των αποτελεσμάτων με σκοπό την συλλογή τελικών συμπερασμάτων.

Κεφάλαιο 1, Κλιματική αλλαγή

Ορισμός Κλιματικής Αλλαγής

Κλιματική αλλαγή ονομάζεται η μακροχρόνια αλλαγή της θερμοκρασίας και των καιρικών φαινομένων. Ως φαινόμενο μπορεί να αναφέρεται τόσο σε μια περιοχή όσο σε όλο τον πλανήτη[12]. Η Γη απορροφά το φως του ήλιου και το εκπέμπει ως θερμότητα. Τα αέρια του θερμοκηπίου που υπάρχουν στην ατμόσφαιρα απορροφούν την θερμότητα που εκπέμπει η Γη και μειώνουν το ρυθμό με τον οποίο αυτή περνάει από την ατμόσφαιρα στο διάστημα[13]. Αυτό έχει ως αποτέλεσμα την αύξηση της μέσης θερμοκρασίας της Γης κατά 0.93 βαθμούς κελσίου τη δεκαετία 2009-2018, και 1.2 βαθμούς το 2020[14].

Η κλιματική αλλαγή καθιστά τα καιρικά φαινόμενα μη προβλέψιμα με αποτέλεσμα να δυσχεραίνεται η καλλιέργεια και η διατήρηση τροφίμων που βασίζονται σε σταθερές θερμοκρασίες και βροχοπτώσεις. Είναι επίσης συνδεδεμένη με άλλες ζημιογόνες καιρικές συνθήκες όπως περισσότερους τυφώνες, πλημμύρες και χειμερινές καταιγίδες. Μια ακόμα επίπτωση της κλιματικής αλλαγής είναι το λιώσιμο των πάγων σε ψυχρές περιοχές, το οποίο ανεβάζει το επίπεδο της θάλασσας με αποτέλεσμα να προκαλούνται ζημιές στις παραθαλάσσιες περιοχές μέσω αύξησης των πλημμυρών και της διάβρωσης του εδάφους.

Επιπτώσεις της Κλιματικής Αλλαγής στην Οικονομία και την Υγεία

Εκτός από το περιβάλλον και το κλίμα, η κλιματική αλλαγή έχει πολύ σοβαρές επιπτώσεις και στους τομείς της οικονομίας και της υγείας.

Όσο αφορά τον τομέα της οικονομίας μια αξιοσημείωτη παρατήρηση είναι ότι το 2017, το κόστος που προκλήθηκε παγκοσμίως από την κλιματική αλλαγή ήταν πάνω από 100 δισεκατομμύρια δολάρια[15]. Περεταίρω οικονομικές έρευνες έδειξαν ότι το μελλοντικό κόστος της κλιματικής αλλαγής θα κυμαίνεται από 2% ως 10% του παγκόσμιου ΑΕΠ[16]. Λόγω αυτών των ενδείξεων, το 2020 το Παγκόσμιο Οικονομικό Φόρουμ ανακήρυξε την κλιματική αλλαγή ως τον σημαντικότερο μακροπρόθεσμο κίνδυνο που αντιμετωπίζει ο κόσμος[17].

Οι επιπτώσεις της κλιματικής αλλαγής στον τομέα τις υγείας οφείλονται σε τραυματισμούς και θανάτους από ακραία καιρικά φαινόμενα[18] καθώς και σε πιο έμμεσα αποτελέσματα της αύξησης της θερμοκρασίας όπως είναι ασθένειες που μεταδίδονται σε πιο ζεστό κλίμα[19]. Ο Παγκόσμιος Οργανισμός υγείας αναμένει αύξηση των ετήσιων θανάτων κατά 250.000 στο διάστημα 2030 με 2050 και κατά 500.000 μετά το 2050[20,21]. Λόγω των παραπάνω ο ΠΟΥ ανακήρυξε την κλιματική αλλαγή ως τη μεγαλύτερη απειλή του 21^{ου} αιώνα[22].

Συνεισφορά της Ενέργειας στην Κλιματική αλλαγή

Η βιομηχανία της ενέργειας είναι ο σημαντικότερος επιβαρυντικός παράγοντας πίσω από την κλιματική αλλαγή. Σε αυτή οφείλεται το 83% της ανθρωπογενούς παραγωγής καθώς και το 65% της συνολικής παραγωγής διοξειδίου του άνθρακα[23]. Η συντριπτική πλειοψηφία αυτής της παραγωγής προέρχεται από τις ανεπτυγμένες χώρες δηλαδή την Ηνωμένες Πολιτείες, την Κίνα και την Ευρωπαϊκή Ένωση. Η λύση για το παραπάνω πρόβλημα είναι η αύξηση της χρήσης των ανανεώσιμων πηγών ενέργειας[24].

Για να αντιμετωπίσει την κλιματική αλλαγή η Ευρωπαϊκή Επιτροπή υπέγραψε την Ευρωπαϊκή Πράσινη Συμφωνία, δηλαδή ένα σύνολο από μέτρα με σκοπό την μείωση της παραγωγής αερίων του θερμοκηπίου[25] κατά τουλάχιστον 50% συγκριτικά με τα επίπεδα του 1990. Η συμφωνία αυτή ήδη έχει εμφανίσει θετικά αποτελέσματα αφού το 2019 η παραγωγή μειώθηκε κατά 25%[26]. Το κύριο μέτρο της ευρωπαϊκής πράσινης συμφωνίας είναι η αύξηση της χρήσης ανανεώσιμων πηγών ενέργειας με στόχο να καλύψει το 32% της συνολικής παραγωγής ενέργειας, σημαντική αύξηση από το 18% που καλύπτει το 2018[27].

Κεφάλαιο 2, Ανανεώσιμες Πηγές Ενέργειας.

Ορισμός

Ανανεώσιμες πηγές ενέργειας είναι αυτές που ανανεώνονται σε ανθρώπινο χρονικό διάστημα, δηλαδή σε χρονικό διάστημα μικρότερο από την ζωή ενός ανθρώπου[28] (σε αντίθεση με τις μη ανανεώσιμες πηγές ενέργειας που μπορεί να χρειαστούν εκατομμύρια χρόνια). Οι κυριότερες μορφές ανανεώσιμης ενέργειας είναι: η ηλιακή, η αιολική, η υδροηλεκτρική και η γεωθερμική.

Η ηλιακή ενέργεια παράγεται από τις ακτίνες που εκπέμπει ο ήλιος στην γη, μέσω ηλιακών φωτοβολταϊκών. Η αιολική ενέργεια παράγεται από τον αέρα με τη χρήση ανεμογεννητριών. Η υδροηλεκτρική ενέργεια συλλέγεται με τη χρήση φραγμάτων που δημιουργούν μια ροή νερού η οποία κινεί μια τουρμπίνα. Τέλος η γεωθερμική ενέργεια παράγεται από την συλλογή της θερμότητας που υπάρχει φυσικά κάτω από την επιφάνεια της γης.

Πλεονεκτήματα και μειονεκτήματα

Οι ανανεώσιμες πηγές ενέργειας έχουν πολλά πλεονεκτήματα[29]. Καταρχάς, όπως προκύπτει και από τον ορισμό τους, δεν πρόκειται ποτέ να τελειώσουν, σε αντίθεση με άλλες μη ανανεώσιμες πηγές όπως το πετρέλαιο. Ως δεύτερο πλεονέκτημα εκ της φύσης τους, οι ανανεώσιμες πηγές ενέργειας παρέχουν ανεξαρτησία, αφού μπορούν να παραχθούν τοπικά και δεν χρειάζονται πρώτη ύλη και καύσιμα τα οποία υπάρχουν σε συγκεκριμένα μέρη του πλανήτη. Ένα ακόμα εξίσου σημαντικό πλεονέκτημα είναι το μικρότερο κόστος διατήρησης και συντήρησης. Επειδή οι μέθοδοι άντλησης της ανανεώσιμης ενέργειας δεν βασίζονται σε καύση, το κόστος που χρειάζονται για την συντήρηση και την

διατήρησή τους είναι αρκετά μικρό. Τέλος οι ανανεώσιμες πηγές ενέργειας προκαλούν πολύ μικρό περιβαλλοντικό κόστος, πράγμα που είναι ίσως το σημαντικότερο πλεονέκτημα τους.

Οι ανανεώσιμες πηγές ενέργειας έχουν και κάποια μειονεκτήματα. Ένα από αυτά είναι το υψηλό αρχικό κόστος επένδυσης. Ενώ, όπως προαναφέρθηκε, το κόστος διατήρησης και συντήρησης των απαραίτητων εξαρτημάτων είναι πιο χαμηλό από αυτό των μη ανανεώσιμων πηγών, το αρχικό κόστος επένδυσης είναι μεγαλύτερο. Το κυριότερο όμως μειονέκτημα είναι η έλλειψη αξιοπιστίας[30]. Η παραγωγή των ανανεώσιμων πηγών ενέργειας εξαρτάται από καιρικά φαινόμενα (παράδειγμα η αιολική ενέργεια απαιτεί ικανοποιητική ταχύτητα αέρα). Αυτό σημαίνει ότι η παραγωγή δεν μπορεί να ρυθμιστεί ώστε να είναι ίση με την κατανάλωση όπως στις μη ανανεώσιμες πηγές (που απλά αυξομειώνεται η ποσότητα του καυσίμου) και επομένως υπάρχει μεγαλύτερη ανάγκη για αποθήκευση της ενέργειας. Τέλος ένα ακόμα μειονέκτημα είναι οι γεωγραφικές συνθήκες της κάθε περιοχής, αφού διαφορετικές περιοχές μπορούν να παράγουν διαφορετικά επίπεδα ενέργειας λόγω των χαρακτηριστικών τους (ηλιοφάνεια, ταχύτητα αέρα, ύπαρξη ποταμιών, βροχοπτώσεις, χιονοπτώσεις κλπ).

Χρήση των Ανανεώσιμων Πηγών Ενέργειας για την Αντιμετώπιση της Κλιματικής Αλλαγής

Οι ανανεώσιμες πηγές ενέργειας έχουν πολύ χαμηλότερη εκπομπή αερίων του θερμοκηπίου όπως το διοξείδιο του άνθρακα. Πιο συγκεκριμένα το φυσικό αέριο παράγει μεταξύ 0.2 με 0.9kg διοξειδίου του άνθρακα για κάθε κιλοβατώρα, η

αιολική ενέργεια 0.1 με 0.18kg και η ηλιακή ενέργεια 0.032 με 0.1kg [31]. Για αυτό το λόγο η Διακυβερνητική Επιτροπή για την Αλλαγή του Κλίματος ανακοίνωσε ότι η αύξηση της χρήσης τους μπορεί να μειώσει την συγκέντρωση αερίων του θερμοκηπίου σε αποδεκτά επίπεδα[32].

Σήμερα βλέπουμε την αύξηση της χρήσης των ανανεώσιμων πηγών ενέργειας κατά περίπου 8% τον χρόνο με αποτέλεσμα ήδη το ένα τρίτο της παγκόσμιας παραγωγής ενέργειας να προέρχεται από ανανεώσιμες πηγές[33]. Πάνω από 30 χώρες καλύπτουν το 20% της ενεργειακής παραγωγής τους με ανανεώσιμες πηγές[34], 47 χώρες καλύπτουν το 50% ενώ 2 χώρες -η Νορβηγία και η Ισλανδία- χρησιμοποιούν αποκλειστικά καθαρή ενέργεια, με πολλές ακόμα χώρες να έχουν θέσει αυτόν τον στόχο[35]. Είναι λοιπόν εμφανές ότι το μέλλον της ενεργειακής παραγωγής είναι οι ανανεώσιμες πηγές ενέργειας.

Κεφάλαιο 3, Ηλεκτρικά αυτοκίνητα

Ηλεκτρικά Αυτοκίνητα και Κλιματική Αλλαγή

Ένας ακόμα τρόπος να καταπολεμηθεί η κλιματική αλλαγή είναι η χρήση ηλεκτρικών αυτοκινήτων. Το 29% της ανθρωπογενούς παραγωγής αερίων του θερμοκηπίου οφείλεται στον τομέα της μεταφοράς, το μεγαλύτερο μέρος του οποίου είναι τα αυτοκίνητα[81]. Τα ηλεκτρικά αυτοκίνητα από την άλλη παράγουν πολύ λιγότερη ποσότητα αερίων του θερμοκηπίου σε σχέση με τα αυτοκίνητα πετρελαίου[45] και επομένως η αύξηση της χρήσης τους είναι απαραίτητη για την αντιμετώπιση της κλιματικής αλλαγής. Ως εκ τούτου πολλές χώρες έχουν λάβει μέτρα με σκοπό την στήριξη της αγοράς ηλεκτρικών αυτοκινήτων, όπως οικονομική ενίσχυση αγοράς, επιδοτήσεις για την ανάπτυξη των απαραίτητων υποδομών καθώς και άλλες ρυθμίσεις με σκοπό την μείωση των αυτοκινήτων πετρελαίου[45].

Είδη Μπαταριών Ηλεκτρικών Αυτοκινήτων

Η μπαταρία είναι το πιο σημαντικό και το πιο ακριβό μέρος ενός ηλεκτρικού αυτοκινήτου. Για να μπορεί να χρησιμοποιηθεί μια μπαταρία σε ένα ηλεκτρικό αυτοκίνητο πρέπει πρώτον να είναι επαναφορτιζόμενη ώστε το αυτοκίνητο να μπορεί να φορτίσει (αφού δεν κάνει χρήση κάποιου καυσίμου) και δεύτερον να έχει μεγάλη χωρητικότητα σε κιλοβατώρες ώστε να προσφέρει αυτονομία πολλών χιλιομέτρων. Η ισχύς των μοντέρνων αυτοκινήτων κυμαίνεται μεταξύ 4.4 και 100 kWh.

Το κυριότερο είδος μπαταριών ηλεκτρικών αυτοκινήτων είναι οι μπαταρίες ιόντων λιθίου (Lithium-ion). Αρχικά σχεδιάστηκαν για laptops και αντίστοιχα ηλεκτρονικά είδη,

αλλά λόγω της μεγάλης τους ενεργειακής χωρητικότητας και διάρκειας ζωής έγινε χρήση τους και στα αυτοκίνητα. Τα κυριότερα μειονεκτήματά τους είναι η μεγάλη εξάρτησή τους από τη θερμοκρασία καθώς και η φθορά που υπόκεινται μετά από εκτεταμένη χρήση. Παρόλα αυτά οι καινούριες παραλλαγές που χρησιμοποιούν τα ηλεκτρικά αυτοκίνητα έχουν πολύ μεγαλύτερη διάρκεια ζωής[41], ενώ εμφανίζουν και υψηλή χωρητικότητα ενέργειας, καθώς και μικρή διάρκεια φόρτισης[42].

Ένα άλλο είδος μπαταρίας είναι οι μπαταρίες μόλυβδου οξέος (Lead-acid). Είναι το φθηνότερο είδος και στο παρελθόν αυτό με τη μεγαλύτερη χρήση. Το σημαντικότερο μειονέκτημά τους είναι η αυξημένη φθορά όταν το επίπεδο φόρτισης πέφτει κάτω από 50% [43]. Οι κύριοι λόγοι της χρήσης τους στο παρελθόν ήταν το μικρό κόστος τους καθώς και η υψηλή τους διαθεσιμότητα. Παρόλα αυτά η χαμηλή σχετική ισχύς τους σε σύγκριση με τα καύσιμα πετρελαίου[46] οδήγησε στη μείωση της χρήσης τους.

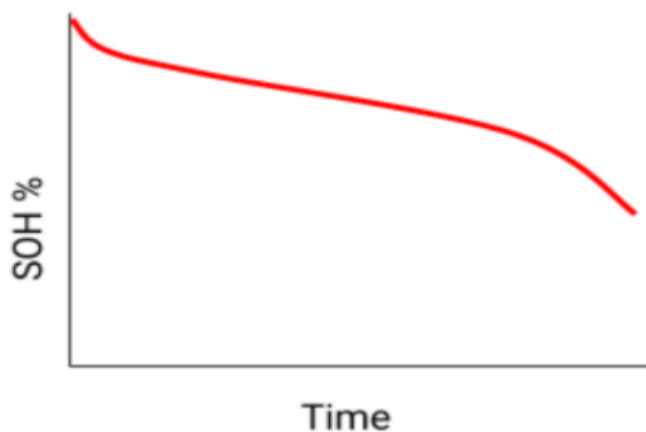
Το τρίτο είδος μπαταριών είναι οι μπαταρίες νικελίου μεταλλικού υβριδίου(Nickel-metal hydride). Χαρακτηρίζονται από πολύ μεγαλύτερη σχετική ισχύ από τις μπαταρίες μόλυβδου οξέος, αλλά έχουν μικρότερη αποδοτικότητα [46]. Ένα από τα κυριότερα πλεονεκτήματά τους είναι η μεγάλη διάρκεια ζωής τους καθώς ορισμένα μοντέλα που κάνουν χρήση αυτού του είδους μπαταρίας μπορούν να χρησιμοποιηθούν για πάνω από 160.000 χιλιόμετρα και πάνω από 10 χρόνια[47].

Τέλος έχουν δει χρήση οι μπαταρίες Ζέμπρα (Zebra) καθώς δεν επηρεάζονται από την εξωτερική θερμοκρασία, αφού απαιτείται σε κάθε περίπτωση να θερμανθούν στους 270 βαθμούς κελσίου για να γίνει κατάλληλη χρήση τους. Αυτό

όμως μπορεί να αποτελέσει σημαντικό μειονέκτημα λόγω της δαπάνης ενέργειας που χρειάζεται για να φτάσει η θέρμανση σε αυτό το επίπεδο.

Φθορά Μπαταριών Ηλεκτρικών Αυτοκινήτων

Ένα από τα μεγαλύτερα προβλήματα που αντιμετωπίζουν τα ηλεκτρικά αυτοκίνητα είναι η φθορά της μπαταρίας τους (Battery Degradation). Φθορά ονομάζεται η σταδιακή μόνιμη μείωση της μέγιστης ενέργειας που μπορεί να αποθηκεύσει ή η ισχύς που μπορεί να αποδώσει η μπαταρία[48]. Βέβαια στα ηλεκτρικά αυτοκίνητα η ισχύς που μπορεί να αποδώσει η μπαταρία είναι πολύ περισσότερη από την απαραίτητη, επομένως μας ενδιαφέρει μόνο η μέγιστη αποθηκευμένη ενέργεια. Ορίζουμε ως κατάσταση υγείας (State Of Health - SOH) το ποσοστό της ενέργειας που μπορεί να αποθηκεύσει η μπαταρία, σε σχέση με το αντίστοιχο ποσό όταν κατασκευάστηκε. Αν για παράδειγμα μια μπαταρία με χωρητικότητα 50 kWh λόγω φθοράς έχει κατάσταση υγείας 85%, τότε εμφανίζει συμπεριφορά σαν μια μπαταρία των 42.5 kWh.



Εικόνα 1: Χρονική μεταβολή της Κατάστασης Υγείας της μπαταρίας ενός ηλεκτρικού αυτοκινήτου

Μια μπαταρία θεωρείται ότι έχει φτάσει στο τέλος της ζωής (End Of Life – EOL) της όταν η κατάσταση υγείας της είναι κάτω από 80%, ή όταν η φυσική της αποφόρτιση είναι πάνω από 5% σε 24 ώρες[49]. Η φθορά της μπαταρίας χωρίζεται σε 2 κατηγορίες δηλαδή σε χρονική γήρανση (Calendar Ageing) και σε κυκλική γήρανση (Cyclic Ageing). Η πρώτη κατηγορία περιγράφει τη φθορά που συμβαίνει όταν η μπαταρία είναι εκτός χρήσης και εξαρτάται από το επίπεδο φόρτισης και τη θερμοκρασία, ενώ η δεύτερη προκύπτει από τη χρήση και εξαρτάται από παράγοντες όπως η θερμοκρασία και ο ρυθμός φόρτισης [51]. Σε αυτή την διπλωματική εργασία θα αναφερθούμε κυρίως στη δεύτερη κατηγορία.

Ο πρώτος παράγοντας που προκαλεί φθορά σε μια μπαταρία ηλεκτρικού αυτοκινήτου είναι η θερμοκρασία. Αυτή μπορεί να μοντελοποιηθεί μαθηματικά με τη σχέση του Arrhenius:

$$r = A * e^{-\frac{E}{kT}}$$

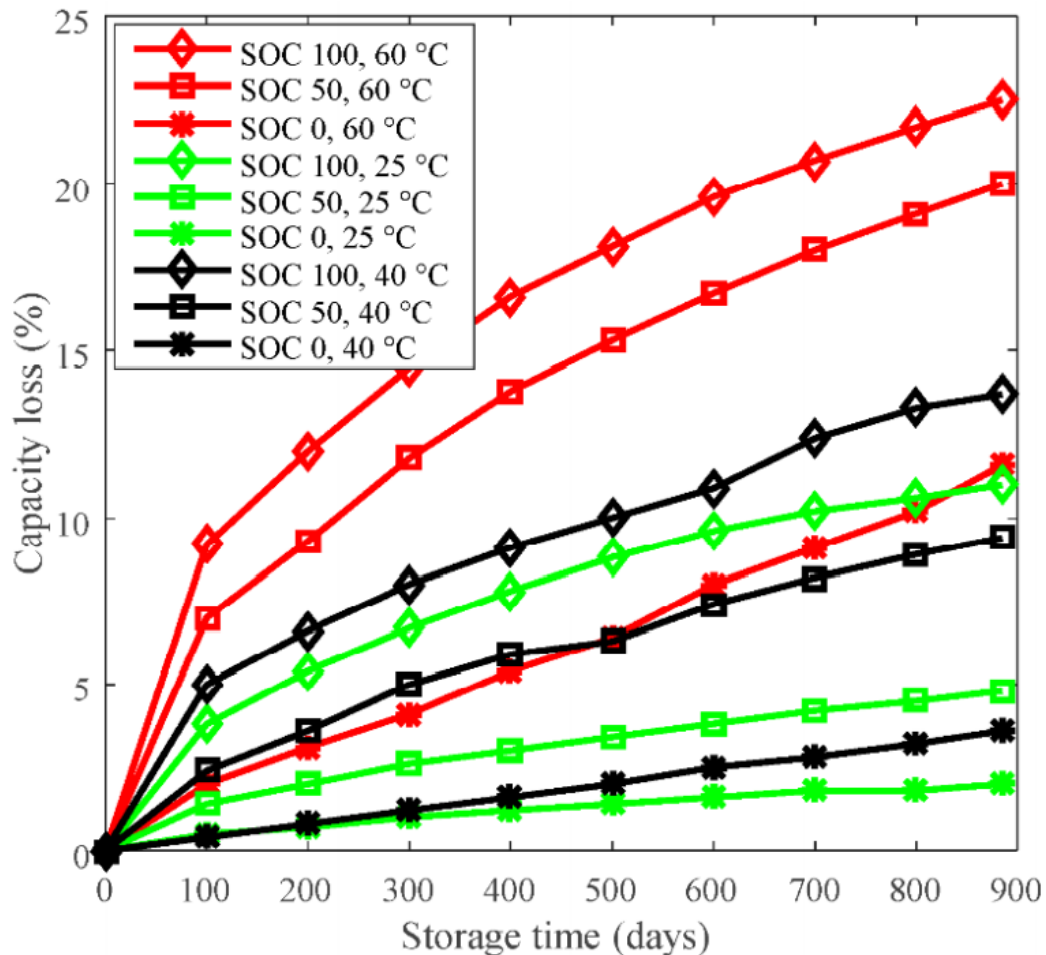
Όπου ως r συμβολίζεται ο ρυθμός της αντίδρασης η οποία ευθύνεται για τη φθορά, E είναι η ενέργεια ενεργοποίησης της αντίδρασης, T είναι η θερμοκρασία της μπαταρίας, k είναι η σταθερά του Boltzmann και A είναι μια σταθερά[52]. Με βάση αυτό το μοντέλο η ζωή της μπαταρίας $L(T)$, η οποία είναι αντιστρόφως ανάλογη του r , μπορεί να γραφτεί ως:

$$L(T) = a * e^{\frac{b}{T}}$$

Όπου a, b είναι σταθερές. Συνεπώς βλέπουμε ότι η ζωή της μπαταρίας μειώνεται όσο αυξάνεται η θερμοκρασία.

Ο δεύτερος παράγοντας που επηρεάζει την μπαταρία είναι το επίπεδο φόρτισης (State Of Charge – SOC). Πειράματα που πραγματοποιήθηκαν από τους Sun, Saxena και Pecht[53],

όπου ηλεκτρικά αυτοκίνητα έμειναν για 900 μέρες εκτός λειτουργίας σε διαφορετικές συνθήκες, δείχνουν ότι όσο αυξάνεται το επίπεδο φόρτισης αυξάνεται και η μείωση της μέγιστης χωρητικότητας.



Εικόνα 2: Μείωση μέγιστης χωρητικότητας μπαταρίας ηλεκτρικών αυτοκινήτων, σε διαφορετικά επίπεδα φόρτισης και θερμοκρασίας, σε διάστημα 900 ημερών

Αυτό οφείλεται στη τάση ανοιχτού κυκλώματος που προκαλείται στην μπαταρία, σε διαφορετικά επίπεδα φόρτισης. Ο πίνακας 1 δείχνει τη τιμή της τάσης ανοιχτού κυκλώματος σε σχέση με το επίπεδο φόρτισης της μπαταρίας για τιμές από 55 έως 100%[51].

SOC (%)	55	60	65	70	75	80	85	90	95	100
Voltage (V)	3.73	3.77	3.82	3.87	3.92	3.98	4.04	4.10	4.16	4.23

Πίνακας 1: Τάση ηλεκτρικής μπαταρίας σε διαφορετικά επίπεδα φόρτισης

Ο τρίτος παράγοντας που μας απασχόλησε σε αυτή τη διπλωματική εργασία είναι ο ρυθμός φόρτισης (C-rate). Πρόκειται για μια αναπαράσταση του ρεύματος φόρτισης, τέτοια ώστε 1C να αποφόρτιζε την μπαταρία σε μια ώρα, ενώ 2C σε 2 ώρες. Ο ρυθμός φόρτισης συμβάλει στην κυκλική γήρανση. Συγκεκριμένα, όσο αυτός αυξάνεται η φθορά που προκαλείται από την κυκλική γήρανση είναι περισσότερη[54]. Μεγαλύτερο ρεύμα φόρτισης οδηγεί επίσης σε υψηλότερη θερμοκρασία η οποία με τη σειρά της αυξάνει την φθορά.

Κεφάλαιο 4, Αγορά Ηλεκτρικής Ενέργειας

Το Ηλεκτρικό Δίκτυο

Το σύστημα ηλεκτρικής ενέργειας αποτελείται από το φυσικό δίκτυο, το οποίο ευθύνεται για την παραγωγή, τη μεταφορά και την κατανομή της ηλεκτρικής ενέργειας, και μια οργανωμένη αγορά ηλεκτρικού ρεύματος[36].

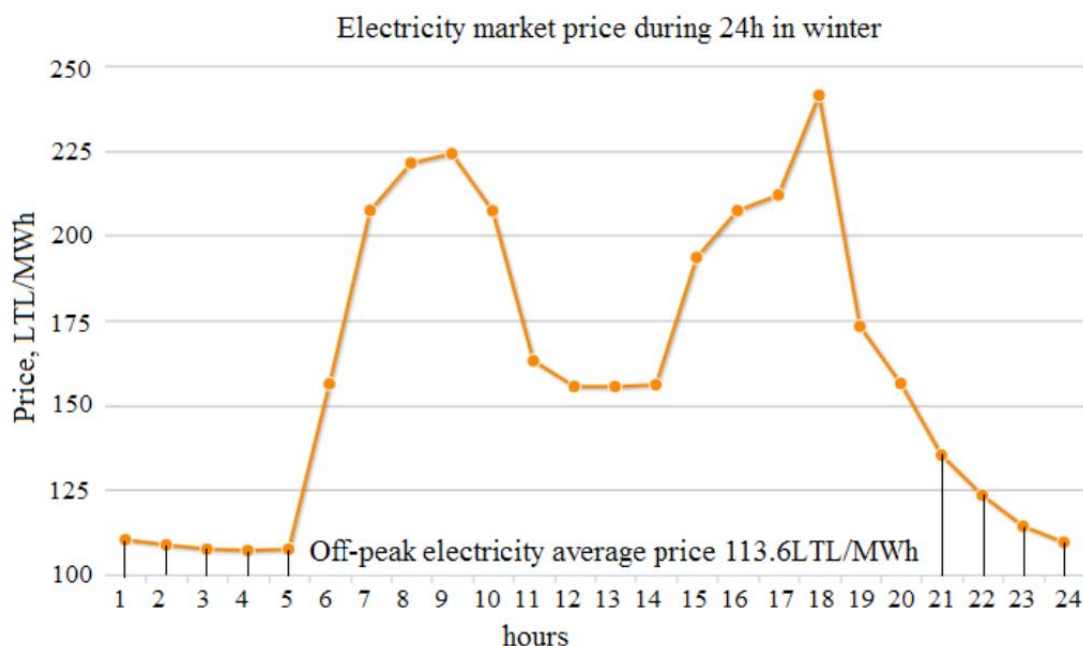
Το ηλεκτρικό δίκτυο συνδέει τους παραγωγούς και τους καταναλωτές ηλεκτρικού ρεύματος με τα δίκτυα μεταφοράς και κατανομής. Έχει δυο βασικές λειτουργίες: να φροντίζει η κατανάλωση ενέργειας να είναι ίση με την παραγωγή της και να αποτρέπει τον έλεγχο της ηλεκτρικής ροής, ώστε οι καταναλωτές να δέχονται ενέργεια από διαφορετικές πηγές. Οι παραγωγοί χωρίζονται σε σταθερής χωρητικότητας (Firm-Capacity), δηλαδή αυτοί που μπορούν να ενεργοποιηθούν και να απενεργοποιηθούν όποτε είναι ανάγκη, και σε μεταβλητής χωρητικότητας (Variable-Capacity), δηλαδή αυτοί που εξαρτώνται από εξωτερικές πηγές (όπως οι καιρικές συνθήκες). Οι περισσότερες ανανεώσιμες πηγές ενέργειας ανήκουν στην δεύτερη κατηγορία.

Η αύξηση της χρήσης των ανανεώσιμων πηγών ενέργειας, που ανήκουν κυρίως στους παραγωγούς μεταβλητής χωρητικότητας, οδηγεί σε μεγαλύτερη ανάγκη ύπαρξης εφεδρικών εργοστασίων ώστε η παραγωγή να μπορεί να ακολουθήσει την κατανάλωση. Αυτή η λύση όμως δεν είναι πολύ ρεαλιστική επειδή οι εταιρείες παραγωγής ηλεκτρικής ενέργειας προσπαθούν να αποφύγουν την συντήρηση εργοστασίων που χρησιμοποιούνται σπάνια. Δύο εναλλακτικές λύσεις -οι οποίες μπορούν να λειτουργήσουν συνεργατικά μεταξύ τους- είναι η κοστολόγηση του ηλεκτρικού ρεύματος με βάση την παραγωγή, με στόχο να μειωθεί η κατανάλωση όταν

η παραγωγή είναι ελλιπής, και η αποθήκευση ηλεκτρικής ενέργειας όταν η παραγωγή είναι περισσότερη από την κατανάλωση, ώστε να γίνεται χρήση της αποθηκευμένης ενέργειας όταν η παραγωγή είναι λιγότερη.

Το Κόστος Ηλεκτρικής Ενέργειας

Το κόστος της ηλεκτρικής ενέργειας δεν είναι σταθερό αλλά αλλάζει με βάση την ώρα. Επομένως όταν η συνολική κατανάλωση είναι υψηλή το κόστος του ρεύματος αυξάνεται, με σκοπό να παροτρύνει τους καταναλωτές να μειώσουν την κατανάλωση τους[37]. Παρακάτω δίνεται ένα παράδειγμα ημερήσιου κόστους ηλεκτρικού ρεύματος[38].



Εικόνα 3: Τιμές κόστους ηλεκτρικού ρεύματος ανά ώρα

Παρατηρούμε ότι το κόστος είναι υψηλότερο το πρωί (κατά τις 7-10) και το μεσημέρι (κατά τις 3-7), ενώ είναι μικρότερο τις πρωινές και τις βραδινές ώρες. Ανάμεσα στους παράγοντες που επηρεάζουν το κόστος είναι τα καιρικά φαινόμενα, η εποχή και η μέρα[98].

Η αύξηση της χρήσης των ανανεώσιμων πηγών ενέργειας -ενώ έχει οδηγήσει στην μείωση του κόστους παραγωγής- προκαλεί την αύξηση της μεταβλητότητας των τιμών του ηλεκτρικού ρεύματος, δηλαδή η διαφορά ανάμεσα στη μικρότερη και τη μεγαλύτερη τιμή σε μια μέρα έχει αυξηθεί[40]. Οι λόγοι για τους οποίους παρατηρείται αυτό είναι οι προαναφερθέντες, δηλαδή η μεταβλητότητα στην παραγωγή που οδηγεί σε μεταβλητότητα στη τιμή κατανάλωσης.

Αυτό σημαίνει ότι η έξυπνη διαχείριση αποθηκευμένου ρεύματος μπορεί να οδηγήσει σε σημαντικά κέρδη, αφού ένας κάτοχος ηλεκτρικών μπαταριών μεγάλης χωρητικότητας μπορεί να αγοράζει και να αποθηκεύει ενέργεια όταν αυτή είναι φθηνή και να πουλάει ενέργεια στο δίκτυο όταν αυτή είναι ακριβή. Μια τέτοια επιλογή, πέρα από κερδοφόρα για τον ιδιώτη, είναι καλή και για το δίκτυο αφού συμβάλει στην προσπάθεια η παραγωγή να φτάσει την κατανάλωση.

Κεφάλαιο 5, Vehicle-to-Grid

Ορισμός

Μια λύση για το πρόβλημα της αποθήκευσης ηλεκτρικής ενέργειας λέγεται vehicle-to-grid. Ο όρος Vehicle-to-grid (V2G) περιγράφει ένα σύστημα στο οποίο ηλεκτρικά αυτοκίνητα επικοινωνούν με το δίκτυο ηλεκτρικής ενέργειας ώστε να προσφέρουν υπηρεσίες, όπως η πώληση ηλεκτρικής ενέργειας στο δίκτυο ή να διαμορφώσουν τον ρυθμό φόρτισής τους. Η ιδέα προτάθηκε από τους Kempton και Latendre το 1996 αλλά μόλις πρόσφατα απέκτησε μεγάλο επιστημονικό ενδιαφέρον λόγω της αύξησης των ηλεκτρικών αυτοκινήτων[55].

Τα ηλεκτρικά οχήματα, όταν δεν χρησιμοποιούνται, είναι συνδεδεμένα στο ηλεκτρικό δίκτυο, για λόγους φόρτισης, αλλά από μόνα τους δεν μπορούν να προσφέρουν ενέργεια και υπηρεσίες στο ηλεκτρικό δίκτυο. Η τεχνολογία Vehicle-to-grid προσφέρει τρόπους στα αυτοκίνητα να επιστρέψουν ενέργεια στο δίκτυο, μέσω ειδικών φορτιστών διπλής κατεύθυνσης, καθώς και τρόπους επικοινωνίας με αυτό. Εισάγει επίσης τον ρόλο του συσσωρευτή (aggregator) ο οποίος αποτελεί έναν μεσάζοντα ανάμεσα στα αυτοκίνητα και στο δίκτυο. Ο συσσωρευτής είναι υπεύθυνος να δίνει οδηγίες στα αυτοκίνητα ως προς τις ενέργειες που θα προβούν, όπως ο ρυθμός φόρτισης ή η πώληση ενέργειας στο δίκτυο, με σκοπό τη μεγιστοποίηση του κέρδους τόσο για το δίκτυο όσο για τους ιδιοκτήτες των αυτοκινήτων[56].

Οι ιδιοκτήτες δίνουν στον συσσωρευτή πληροφορίες σχετικά με το χρόνο που θα είναι συνδεδεμένοι στο δίκτυο, το επίπεδο φόρτισης που χρειάζονται κατά την αναχώρησή τους καθώς και πληροφορίες σχετικά με την κατάσταση της μπαταρίας τους. Ο συσσωρευτής βλέπει τις αντίστοιχες

πληροφορίες πολλών αυτοκινήτων, το κόστος της ηλεκτρικής ενέργειας (καθώς και προβλέψεις για το αντίστοιχο κόστος στο σύντομο μέλλον) και τις ανάγκες του ηλεκτρικού δικτύου. Στη συνέχεια απαντάει στα αυτοκίνητα με τον βέλτιστο τρόπο φόρτισης και αποφόρτισης ώστε να μεγιστοποιηθεί το κέρδος που μπορεί να διανεμηθεί στο κάθε αυτοκίνητο με το να παρέχει υπηρεσίες στο δίκτυο ενώ φροντίζει και να τηρηθούν οι περιορισμοί που προκύπτουν από αυτό.

Πλεονεκτήματα του Vehicle-to-Grid

Η τεχνολογία Vehicle-to-Grid έχει πολλά πλεονεκτήματα. Καταρχάς ένα ελεγχόμενο σύστημα φόρτισης μπορεί να φορτίσει ένα αυτοκίνητο με μικρότερο κόστος δίνοντας ρεύμα με μεγαλύτερο ρυθμό όταν αυτό είναι φθηνό και με μικρότερο όταν αυτό είναι ακριβό. Έτσι ο ιδιοκτήτης ενός ηλεκτρικού αυτοκινήτου μπορεί να μειώσει το κόστος της χρήσης του ακόμα και αν αρνηθεί να προσφέρει υπηρεσίες στο ηλεκτρικό δίκτυο.

Το μεγαλύτερο κέρδος όμως μπορεί να επιτευχθεί όταν τα ηλεκτρικά αυτοκίνητα προσφέρουν υπηρεσίες στο ηλεκτρικό δίκτυο. Μια υπηρεσία που μπορεί να προσφέρει στο δίκτυο ένα σύστημα V2G είναι η ρύθμιση της τάσης και της συχνότητας του ηλεκτρικού ρεύματος[57]. Σε ένα δίκτυο ηλεκτρικής ενέργειας η ρύθμιση της συχνότητας έχει σκοπό την ισορρόπηση της κατανάλωσης και της παραγωγής της ενεργούς ισχύος, ενώ η ρύθμιση της τάσης ισορροπεί την άεργο ισχύ[58]. Ένα σύστημα V2G μπορεί να ρυθμίσει την συχνότητα λόγω του γρήγορου ρυθμού φόρτισης και αποφόρτισης των μπαταριών των ηλεκτρικών αυτοκινήτων. Μπορεί επίσης να ρυθμίσει το επίπεδο της τάσης παρέχοντας

ενεργό η άεργο ισχύ, πράγμα που μπορεί να επιτευχθεί με την κατάλληλη επιλογή της γωνίας της φάσης του ρεύματος που προσφέρει ο κάθε φορτιστής. Ένα αυτοκίνητο που παρέχει την παραπάνω υπηρεσία σε ένα σύστημα V2G μπορεί, ανάλογα με την περιοχή, να επιφέρει μέχρι και 5.000\$ κέρδος τον χρόνο[60].

Η δεύτερη υπηρεσία που μπορεί να προσφέρει ένα σύστημα V2G είναι η μετατόπιση φορτίου (Load Shifting). Μετατόπιση φορτίου, γνωστή και ως διαχείριση ζήτησης ενέργειας (Demand-side Management) είναι η ισορρόπηση της ενεργειακής κατανάλωσης με την παροχή μέσω του ελέγχου του φορτίου αντί για την παραγωγή των ηλεκτρικών εργοστασίων[61]. Ένα σύστημα V2G μπορεί μέσω έξυπνης φόρτισης να φορτίσει τα αυτοκίνητα όταν η ζήτηση είναι χαμηλή και να τα αποφορτίσει όταν είναι ψηλή. Με αυτόν τον τρόπο μειώνεται σημαντικά η επίδραση των ηλεκτρικών αυτοκινήτων στο ηλεκτρικό δίκτυο[59].

Τέλος, ίσως το μεγαλύτερο πλεονέκτημα της τεχνολογίας V2G, είναι ο αποθηκευτικός χώρος που παρέχει, σε συνδυασμό με τις ανανεώσιμες πηγές ενέργειας. Όπως αναφέρθηκε πιο πάνω οι ανανεώσιμες πηγές ενέργειας πάσχουν από έλλειψη προσαρμοστικότητας, δηλαδή δεν είναι εύκολο να ρυθμιστεί η παραγωγή ώστε να φτάσει την κατανάλωση. Ένα σύστημα V2G με πολλά αυτοκίνητα μπορεί να αξιοποιηθεί σαν αποθήκη ηλεκτρικής ενέργειας υψηλής χωρητικότητας. Επομένως τα ηλεκτρικά αυτοκίνητα μπορούν να φορτίσουν και να αποφορτίσουν τις μπαταρίες τους ανάλογα με τις ανάγκες του ηλεκτρικού δικτύου, καθώς επίσης και να αποθηκεύσουν περίσσια ενέργεια είτε για δική τους χρήση είτε για να την παρέχουν στο δίκτυο μετέπειτα[62]. Με αυτό το τρόπο η τεχνολογία vehicle-to-grid προσφέρει μια

λύση σε ένα από τα σημαντικότερα προβλήματα της σύγχρονης επιστήμης.

Συμπερασματικά τα συστήματα vehicle-to-grid μπορούν να προσφέρουν ιδιαίτερο κέρδος τόσο στους ιδιοκτήτες των ηλεκτρικών αυτοκινήτων όσο και στο ηλεκτρικό δίκτυο, ενώ βοηθούν στην ανάπτυξη μιας πιο πράσινης κοινωνίας.

Μειονεκτήματα και Προκλήσεις του Vehicle-to-Grid

Η τεχνολογία Vehicle-to-Grid, όπως και κάθε τεχνολογία, έχει ορισμένα μειονεκτήματα καθώς και κάποιες προκλήσεις που πρέπει να αντιμετωπιστούν για την αποτελεσματικότερη αξιοποίησή της. Το μεγαλύτερο μειονέκτημα είναι η φθορά που πιθανώς προκαλείται στις μπαταρίες των ηλεκτρικών αυτοκινήτων, όταν αυτά παίρνουν μέρος σε ένα σύστημα V2G. Όταν ένα ηλεκτρικό αυτοκίνητο παρέχει στο δίκτυο ηλεκτρικής ενέργειας υπηρεσίες είναι πιθανό το ρεύμα με το οποίο φορτίζεται ή αποφορτίζεται να έχει υψηλή τιμή ή μεγάλη γωνία φάσης. Σε τέτοιες περιπτώσεις προκαλείται αυξημένη φθορά στην μπαταρία του αυτοκινήτου, όπως συζητήθηκε παραπάνω. Η αποθήκευση ηλεκτρικού ρεύματος και η μετατόπιση φορτίου οδηγούν επιπλέον σε μεγαλύτερο επίπεδο φόρτισης, πράγμα που μειώνει ακόμα περισσότερο τον χρόνο ζωής της μπαταρίας. Αυτό το πρόβλημα μπορεί να αντιμετωπιστεί μερικώς με κατάλληλο έλεγχο και αποφάσεις κατά την φόρτιση των ηλεκτρικών αυτοκινήτων[63].

Το δεύτερο σημαντικό μειονέκτημα των συστημάτων V2G είναι η φθορά που μπορεί να προκαλέσουν στο ηλεκτρικό δίκτυο. Τα Συστήματα V2G αυξάνουν τις απώλειες των μετασχηματιστών διανομής, την απόκλιση της τάσης και την αρμονική παραμόρφωση[65]. Αυτό έχει ως πιθανό

αποτέλεσμα την φθορά των μετασχηματιστών διανομής[64] και άρα δημιουργεί την ανάγκη για μεγαλύτερη επένδυση στο δίκτυο διανομής, τόσο στα καλώδια και στις γραμμές όσο και στους καταχωρητές των μετασχηματιστών. Παρόλα αυτά το πρόβλημα της φθοράς του ηλεκτρικού δικτύου μπορεί να αντιμετωπιστεί σε σημαντικό βαθμό με την χρήση ελεγχόμενων μεθόδων φόρτισης[66].

Τέλος μια μεγάλη πρόκληση με την οποία βρίσκεται αντιμέτωπη η τεχνολογία V2G είναι ο ανθρώπινος παράγοντας. Για να μπορέσει να εφαρμοστεί το V2G σε ένα σύστημα χρειάζεται ένας ικανοποιητικά μεγάλος στόλος ηλεκτρικών αυτοκινήτων. Αν και το πλήθος των ηλεκτρικών αυτοκινήτων είναι πολύ πιθανό να αυξηθεί, δεν είναι απόλυτα σίγουρο ότι οι ιδιοκτήτες των αυτοκινήτων θα θέλουν να συνεισφέρουν στο σύστημα V2G. Επομένως θα χρειαστεί η εφαρμογή κατάλληλων μέτρων πειθούς όπως χρηματικές επιβραβεύσεις ή επιχορηγήσεις για να αντιμετωπιστεί αυτός ο παράγοντας. Είναι πιθανό επίσης να υπάρξει αντίσταση από ορισμένες βιομηχανίες (όπως για παράδειγμα αυτή του πετρελαίου) αφού θα μειωθούν κατά πολύ τα κέρδη τους μιας και η τεχνολογία V2G καθιστά την αύξηση της χρήσης των ανανεώσιμων πηγών ενέργειας πιο πιθανή.

Vehicle-to-Grid και Τεχνητή Νοημοσύνη

Όλες οι λειτουργίες του V2G στηρίζονται πάνω σε αποφάσεις σχετικά με τη φόρτιση. Τόσο για να προσφέρει ένα αυτοκίνητο υπηρεσίες στο δίκτυο όσο και για την πιο αποδοτική φόρτισή του (δηλαδή αυτή που προκαλεί λιγότερη φθορά στην μπαταρία ή που δίνει το μικρότερο κόστος) πρέπει να παρθεί μια απόφαση σχετικά με τα χαρακτηριστικά του

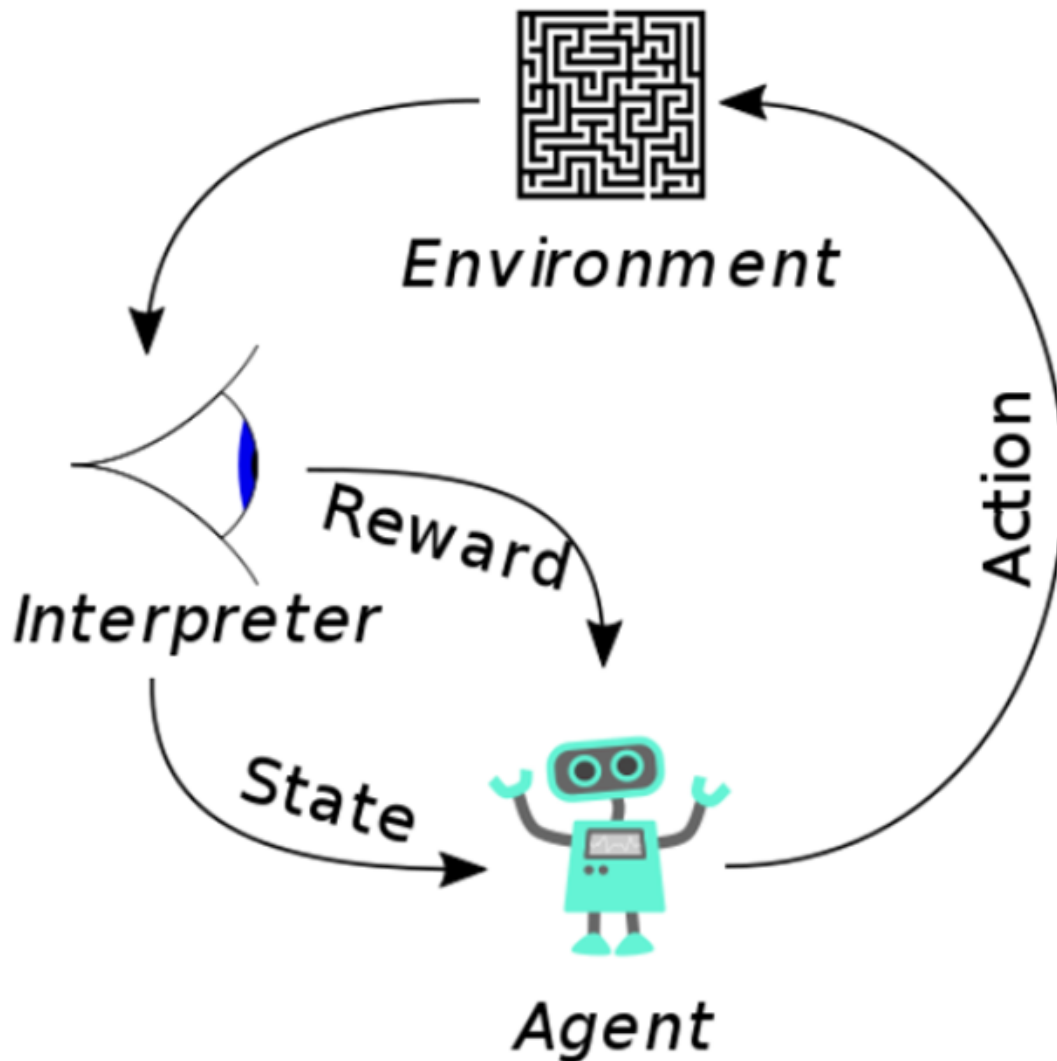
ρεύματος φόρτισης και αποφόρτισης. Για την καλύτερη οργάνωση και χρησιμοποίηση των συστημάτων V2G είναι απαραίτητη η ύπαρξη του συσσωρευτή ο οποίος βλέπει μεγάλο πλήθος ηλεκτρικών αυτοκινήτων καθώς και τις ανάγκες του δικτύου και με βάση αυτές τις πληροφορίες παίρνει αποφάσεις για το κάθε αυτοκίνητο.

Αυτό που καλείται να επιλύσει ο συσσωρευτής είναι ένα πρόβλημα βελτιστοποίησης δηλαδή πως θα βγάλει το μέγιστο κέρδος με βάση τα δεδομένα που του δίνονται ενώ πρέπει να τηρήσει ορισμένους περιορισμούς. Πρέπει επίσης, για την εύρεση των βέλτιστων αποτελεσμάτων, να λύσει και κάποια προβλήματα πρόβλεψης. Ορισμένα παραδείγματα τέτοιων προβλημάτων είναι: οι αφίξεις μελλοντικών αυτοκινήτων, οι τιμές του ρεύματος στο προσεχές μέλλον και οι ανάγκες που μπορεί να αντιμετωπίσει σύντομα το δίκτυο ηλεκτρικής ενέργειας. Και για τα δύο είδη προβλημάτων ένας πολύ καλός τρόπος επίλυσής τους είναι η χρήση μεθόδων τεχνητής νοημοσύνης και μηχανικής μάθησης. Στην συνέχεια αυτής της διπλωματικής εργασίας θα γίνει χρήση ενισχυμένης μάθησης (Reinforcement Learning) για τη διαδικασία απόφασης ενός συστήματος V2G.

Κεφάλαιο 6, Ενισχυτική Μάθηση

Ορισμός

Ενισχυτική μάθηση (Reinforcement Learning) είναι μια από τις 3 βασικές κατηγορίες μηχανικής μάθησης, μαζί με την επιβλεπόμενη μάθηση (Supervised Learning) και τη μη επιβλεπόμενη μάθηση (Unsupervised Learning). Σε ένα περιβάλλον ενισχυτικής μάθησης ένας έξυπνος πράκτορας (agent) παίρνει αποφάσεις (actions) με βάση τις παρατηρήσεις (observations) ενός περιβάλλοντος (environment) με σκοπό τη μεγιστοποίηση ενός αθροιστικού κέρδους (reward).



Εικόνα 4: Μοντέλο Ενισχυτικής μάθησης

Οι βασικότερες διαφορές της με την επιβλεπόμενη μάθηση είναι ότι πρώτον δεν χρειάζεται ζευγάρια βέλτιστης εισόδου και εξόδου ως δεδομένα και δεύτερον οι μη βέλτιστες αποφάσεις δεν χρειάζεται να διορθωθούν. Αντίθετα ο σκοπός είναι να βρεθεί η βέλτιστη ισορροπία ανάμεσα στην εξερεύνηση νέων αποφάσεων και την εκμετάλλευση της υπάρχουσας γνώσης πάνω στο σύστημα.

Σε ένα πρόβλημα ενισχυτικής μάθησης το περιβάλλον ορίζεται ως μια διαδικασία απόφασης Markov (Markov Decision Process – MDP) και περιέχει τα παρακάτω σύνολα:

- Εποχές απόφασης (Decision Epochs) T . Πρόκειται για ένα σύνολο που μπορεί να είναι ή άπειρο ή πεπερασμένο ενώ οι τιμές που περιλαμβάνει μπορεί να είναι ή διακριτές ή συνεχείς. Τα στοιχεία που ανήκουν στο T συμβολίζονται ως t .
- Χώρος καταστάσεων (States Space) S . Πρόκειται για το σύνολο καταστάσεων στο οποίο μπορεί να βρεθεί ένας πράκτορας και να κληθεί να πάρει απόφαση πάνω σε αυτές.
- Χώρος αποφάσεων (Action Space) A . Περιέχει τις πιθανές αποφάσεις που μπορεί να πάρει ο πράκτορας. Σε κάθε εποχή απόφασης ο πράκτορας καλείται να πάρει μια απόφαση ώστε να μεταβεί από τη μια κατάσταση στην επόμενη και να λάβει την αντίστοιχη επιβράβευση.
- Επιβραβεύσεις (Rewards) R . Μια επιβράβευση είναι μια συνάρτηση που με βάση την εποχή απόφασης, την κατάσταση του πράκτορα και την απόφασή του έχει ως έξοδο έναν αριθμό, δηλαδή την αξία της συγκεκριμένης απόφασης. Αναπαρίσταται ως $r_t(s, a)$. Ο σκοπός της είναι να υποδείξει στον πράκτορα ποιος στόχος πρέπει

να επιτευχθεί (πράγμα που δεν είναι ίδιο με το πως αυτός στόχος πρέπει να επιτευχθεί).

- Κατανομή πιθανοτήτων μεταβάσεων κατάστασης (State-transition probabilities) p . Πρόκειται για την πιθανότητα να παρθεί μια απόφαση από μια κατάσταση και να γίνει μετάβαση σε μια συγκεκριμένη επόμενη κατάσταση.

Μαθηματικό υπόβαθρο

Πολιτική (policy) είναι ο τρόπος με τον οποίο ένας πράκτορας επιλέγει μια απόφαση σε μια δεδομένη κατάσταση. Πρόκειται για μια συνάρτηση $\pi: A \times S \rightarrow [0,1]$ και συμβολίζει την πιθανότητα ο πράκτορας να πάρει μια απόφαση a όταν βρίσκεται σε κατάσταση s :

$$\pi(a, s) = \Pr(a_t = a | s_t = s)$$

Συχνά ο στόχος ενός συστήματος ενισχυτικής μάθησης είναι να βρεθεί η βέλτιστη πολιτική, δηλαδή το βέλτιστο σύνολο αποφάσεων. Μια πολιτική μπορεί να είναι τυχαία ή ντετερμινιστική, δηλαδή να επιστρέφει πιθανότητες πάνω στο σύνολο αποφάσεων ή απλά μια σίγουρη απόφαση. Μπορεί επίσης να είναι ιστορικά εξαρτώμενη ή Μαρκοβιανή, δηλαδή να είναι συνάρτηση των προηγούμενων καταστάσεων και αποφάσεων ή να εξαρτάται μόνο από την τωρινή κατάσταση.

Συνάρτηση Αξίας (Value function) ορίζεται ως το αναμενόμενο κέρδος μιας κατάστασης s όταν ακολουθείται η πολιτική π . Με άλλα λόγια πρόκειται για μια εκτίμηση του πόσο καλό είναι να βρίσκεται ο πράκτορας σε αυτή τη κατάσταση

$$V_T^\pi(s) = E[R] = E \left[\sum_{t=0}^T \gamma^t r_t | s_0 = s \right]$$

Όπου το R συμβολίζει το κέρδος και ορίζεται ως το άθροισμα των μελλοντικών επιβραβεύσεων.

$$R = \sum_{t=0}^T \gamma^t r_t$$

Όπου r_t είναι η επιβράβευση την εποχή t και $\gamma \in (0,1]$ είναι ο ρυθμός έκπτωσης και έχει ως σκοπό οι αποφάσεις στο μακρινό μέλλον να έχουν μικρότερη επίπτωση από αυτές στο σύντομο. Μπορούμε να ορίσουμε την συνάρτηση μέγιστης αξίας ως εξής:

$$V_T^*(s) = \sup_{\pi \in \Pi} V_T^\pi(s)$$

Δηλαδή ως το μέγιστο δυνατό κέρδος όταν η αρχική κατάσταση είναι s . Μπορεί να οριστεί επίσης η συνάρτηση:

$$u_t^\pi(s) = E \left[\sum_{i=t}^T \gamma^i r_i(s_i^\pi, a_i^\pi) \mid s_t^\pi = s \right].$$
 Προφανώς ισχύει ότι

$$V_T^\pi(s) = u_1^\pi(s). \text{ Ορίζεται επίσης η } u_t^*(s) = \sup_{\pi \in \Pi} u_t^\pi(s)$$
 η οποία

συμβολίζει το μέγιστο δυνατό μέσο εκτιμώμενο κέρδος που μπορεί να αποδοθεί από οποιαδήποτε πολιτική, αν την στιγμή t βρισκόμαστε στην κατάσταση s .

Ορίζεται ως συνάρτηση αξίας κατάστασης-απόφασης (state-action value function), το αναμενόμενο κέρδος μιας πολιτικής όταν βρίσκεται στην κατάσταση s_t και επιλέγει την απόφαση a_t .

$$Q_t^\pi(s_t, a) = r_t(s_t, a) + \sum_{s \in S} p_t(s \mid s_t, a) u_t^\pi(s) =$$

$$r_t(s_t, a) + \sum_{s \in S} p_t(s \mid s_t, a) \max_{a'} Q_{t+1}^\pi(s, a')$$

Όπου p_t η συνάρτηση μετάβασης.

Τέλος μπορούμε να ορίσουμε την συνάρτηση πλεονεκτήματος (Advantage function) $A(s,a)$ ως:

$$A(s, a) = Q(s, a) - V(s)$$

Για να υπολογίσουμε την βέλτιστη πολιτική χρησιμοποιούμε τις εξισώσεις Bellman. Έτσι η $u_t^*(s)$ μπορεί να προσεγγισθεί μέσω της $u_t^{Bellman}(s)$ η οποία συμβολίζεται $u_t^B(s)$. Οι εξισώσεις του Bellman είναι οι εξής:

1. $u_t^B(s_t) = \max_a r_T(s_T, a) \quad \forall s_T$ δηλαδή η τιμή της συνάρτησης u για κάθε τερματική κατάσταση είναι η μέγιστη ανταμοιβή της.

2. $\forall t \in \{T - 1, T - 2, \dots, 1\}$ και $\forall s_t$:

$$u_t^B(s_t) = \max_{a \in A} (Q_t(s_t, a))$$

3. Η βέλτιστη πολιτική $\pi_t(s_t)$ υπολογίζεται με την ανάθεση $Q_t(s_t, \pi_t(s_t)) = \max_{a \in A} Q_t(s_t, a)$ δηλαδή $\pi_t(s_t) = \operatorname{argmax}_{a \in A} Q_t(s_t, a)$

Αποδεικνύεται ότι αφού υπολογιστεί το u^B με μεθόδους δυναμικού προγραμματισμού ισχύει $u^* = u^B$.

Ο υπολογισμός του u^B με τις εξισώσεις του Bellman έχει πολυπλοκότητα $\Theta(S^2AT)$ που είναι κατά πολύ βελτιωμένη σε σχέση με την εξαντλητική μέθοδο η οποία θα χρειαζόταν $\Theta(A^{ST}TK^2)$ υπολογιστικά βήματα. Παρόλα αυτά ο ακριβής υπολογισμός της λύσης των εξισώσεων συνήθως δεν είναι εφικτός καθώς το σύνολο των καταστάσεων είναι πολύ μεγάλο, και σε πολλές περιπτώσεις μπορεί και να απειρίζεται. Για αυτό το λόγο γίνεται χρήση αλγορίθμων τεχνητής

νοημοσύνης και νευρωνικών δικτύων με σκοπό την προσέγγιση του u^B .

Αλγόριθμοι Ενισχυτικής Μάθησης.

DQN

Ο αλγόριθμος Q-learning είναι ένας αλγόριθμος χωρίς μοντέλο ο οποίος μαθαίνει την αξία μιας απόφασης σε μια συγκεκριμένη κατάσταση. Δεν χρειάζεται να γνωρίζει το μοντέλο του συστήματος (για αυτό και λέγεται χωρίς μοντέλο – model-free) και για αυτό μπορεί να χρησιμοποιηθεί σε πολλά προβλήματα με στοχαστικές μεταβάσεις. Μπορεί να υπολογίζει την βέλτιστη πολιτική για οποιαδήποτε πεπερασμένη διαδικασία απόφασης Markov μέσω της μεγιστοποίησης της αναμενόμενης αξίας της συνολικής επιβράβευσης για κάθε βήμα ξεκινώντας από μια αρχική κατάσταση[67]. Το όνομά του προκύπτει από την συνάρτηση που υπολογίζει, δηλαδή την συνάρτηση αξίας κατάστασης-απόφασης (Q-function).

Ο αλγόριθμος είναι αρκετά απλός. Πρώτα αρχικοποιεί έναν πίνακα $Q(s, a)$ με τυχαίες τιμές. Ύστερα, για κάθε βήμα, ξεκινώντας από μια κατάσταση s , επιλέγει μια απόφαση a_t και μεταβαίνει στην αντίστοιχη επόμενη κατάσταση s' . Για να μειωθεί η εξάρτηση από την αρχική τυχαία αρχικοποίηση η επιλογή απόφασης γίνεται με μια μέθοδο η οποία ονομάζεται epsilon – greedy. Ο αλγόριθμος επιλέγει μια τυχαία απόφαση με πιθανότητα ϵ ή την απόφαση που προκύπτει από τον πίνακα Q με πιθανότητα $1-\epsilon$. Η τιμή του ϵ αρχίζει από το 1 και μειώνεται με σταθερό ρυθμό. Έτσι στην αρχή του αλγορίθμου, όταν αυτός δεν έχει αποκτήσει αρκετή γνώση για το σύστημα, γίνεται περισσότερη εξερεύνηση στο πεδίο των λύσεων ενώ

στην συνέχεια γίνεται καλύτερη χρησιμοποίηση της υπάρχουσας γνώσης. Τέλος αφού παρθεί η απόφαση a_t ο αλγόριθμος ενημερώνει την τιμή $Q(s,a)$ ως εξής:

$$Q(s, a_t) = (1 - \alpha)Q(s, a_t) + \alpha(r + \gamma \max_{a'} Q(s', a') - Q(s, a_t))$$

Όπου α είναι ο ρυθμός μάθησης (learning rate), r είναι η επιβράβευση και γ είναι ο συντελεστής έκπτωσης. Ο όρος $\max_{a'} Q(s', a')$ είναι μια εκτίμηση για την αξία της επόμενης κατάστασης και προκύπτει από τον πίνακα Q . Ο αλγόριθμος σε ψευδοκώδικα είναι ο εξής:

```
Initialize  $Q(s, a)$  arbitrarily
Repeat (for each episode):
  Initialize  $s$ 
  Repeat (for each step of episode):
    Choose  $a$  from  $s$  using policy derived from  $Q$ 
    (e.g.,  $\epsilon$ -greedy)
    Take action  $a$ , observe  $r, s'$ 
     $Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$ 
     $s \leftarrow s'$ 
  until  $s$  is terminal
```

Εικόνα 5: Q-learning algorithm [68]

Οι κύριες υπερπαραμέτροι του αλγορίθμου Q-learning είναι ο ρυθμός μάθησης, ο συντελεστής έκπτωσης και οι αρχικές συνθήκες. Ο ρυθμός μάθησης καθορίζει κατά πόσο καινούρια πληροφορία αντικαθιστά την παλιά. Σε πλήρως ντετερμινιστικά περιβάλλοντα ο ρυθμός 1 είναι βέλτιστος. Σε στοχαστικά περιβάλλοντα όμως μικρότερες τιμές δίνουν μεγαλύτερη πιθανότητα σύγκλισης. Η πιο συνηθισμένη τιμή είναι $\alpha=0.1$.

Ο συντελεστής έκπτωσης καθορίζει την σημασία μελλοντικών επιβραβεύσεων. Αν αυτός τεθεί στο 0 ο πράκτορας ενδιαφέρεται μόνο για την άμεση επιβράβευση ενώ όταν η τιμή είναι 1 υπολογίζει και μελλοντικές αποφάσεις. Αν η τιμή του ξεπεράσει το 1 ο αλγόριθμος συνήθως αποκλίνει.

Στην περίπτωση που το πλήθος αποφάσεων σε μια εποχή δεν είναι σταθερό, μεγάλες τιμές του συντελεστή έκπτωσης οδηγούν σε εποχές μεγαλύτερης διάρκειας. Σε πολλές περιπτώσεις η βέλτιστη πρακτική είναι ο συντελεστής να αρχικοποιηθεί με κάποια μικρή τιμή η οποία να αυξηθεί σταδιακά κατά την διαδικασία μάθησης[70].

Η αρχική κατάσταση του πίνακα Q έχει και αυτή κάποια σημασία. Αν και ο πίνακας Q στην αρχή αποκτάει τυχαίες τιμές είναι καλύτερα αυτές να είναι ψηλές επειδή τέτοια αρχικοποίηση επιβραβεύει την εξερεύνηση[71]. Πολλές φορές η πρώτη απόφαση αντικαθιστά πλήρως την τυχαία αρχική κατάσταση αντί να την ενημερώσει, πράγμα που μπορεί να επιφέρει γρηγορότερη μάθηση.

Ο αλγόριθμος Q-learning έχει πολύ καλά αποτελέσματα στην θεωρία αλλά στην πράξη συνήθως δεν είναι εφικτός. Για να αποδώσει χρειάζεται περιορισμένο σύνολο καταστάσεων το οποίο μπορεί να αποθηκευτεί ολόκληρο στην μνήμη. Στις περισσότερες περιπτώσεις το σύνολο καταστάσεων είναι συνεχές και σε ορισμένες μπορεί και να απειρίζεται. Για αυτό στην πράξη χρησιμοποιούμε τον αλγόριθμο DQN (Deep Q-Learning). Ο αλγόριθμος DQN αντικαθιστά τον πίνακα Q με ένα replay buffer και ένα νευρωνικό δίκτυο. Το νευρωνικό δίκτυο έχει ως είσοδο μια κατάσταση s και ως έξοδο μια απόφαση a . Με κάθε απόφαση a , σε κάποια κατάσταση s , αποθηκεύουμε στο replay buffer την μετάβαση $\langle s, a, r, s' \rangle$ όπου r η επιβράβευση και s' η επόμενη κατάσταση. Αυτή την πληροφορία τη χρησιμοποιούμε για την εκπαίδευση του νευρωνικού δικτύου με τον κατάλληλο ορισμό της συνάρτησης απώλειας (Loss function) ως εξής:

$$Loss = L(\theta) = \left(r + \gamma \max_{a'} Q(s', a'; \theta') - Q(s, a; \theta) \right)^2$$

Όπου θ τα βάρη του νευρωνικού. Η εκπαίδευση του νευρωνικού γίνεται με την μέθοδο Gradient Decent όπως και στην επιβλεπόμενη μάθηση. Για την καλύτερη σύγκλιση του αλγορίθμου χρησιμοποιούνται 2 ίδια νευρωνικά δίκτυα. Το πρώτο είναι αυτό που παίρνει τις αποφάσεις και το δεύτερο αυτό που εκπαιδεύεται. Μετά από έναν αριθμό βημάτων εκπαίδευσης το πρώτο γίνεται ένα αντίγραφο του δεύτερου και συνεχίζεται η διαδικασία. Έτσι αποφεύγεται η ταλάντωση που θα προκαλούνταν από τη συνεχή εκπαίδευση στο δίκτυο απόφασης. Ο αλγόριθμος DQN σε ψευδοκώδικα είναι ο εξής:

Algorithm: deep Q-learning with experience replay.

Initialize replay memory D to capacity N

Initialize action-value function Q with random weights θ

Initialize target action-value function \hat{Q} with weights $\theta^- = \theta$

For episode = 1, M **do**

Initialize sequence $s_1 = \{x_1\}$ and preprocessed sequence $\phi_1 = \phi(s_1)$

For $t = 1, T$ **do**

With probability ε select a random action a_t

otherwise select $a_t = \operatorname{argmax}_a Q(\phi(s_t), a; \theta)$

Execute action a_t in emulator and observe reward r_t and image x_{t+1}

Set $s_{t+1} = s_t, a_t, x_{t+1}$ and preprocess $\phi_{t+1} = \phi(s_{t+1})$

Store transition $(\phi_t, a_t, r_t, \phi_{t+1})$ in D

Sample random minibatch of transitions $(\phi_j, a_j, r_j, \phi_{j+1})$ from D

Set $y_j = \begin{cases} r_j & \text{if episode terminates at step } j+1 \\ r_j + \gamma \max_{a'} \hat{Q}(\phi_{j+1}, a'; \theta^-) & \text{otherwise} \end{cases}$

Perform a gradient descent step on $(y_j - Q(\phi_j, a_j; \theta))^2$ with respect to the network parameters θ

Every C steps reset $\hat{Q} = Q$

End For

End For

Εικόνα 6: DQN algorithm[72]

IQL

Σε πολλές εφαρμογές ενισχυτικής μάθησης είναι απαραίτητο να παρθούν παραπάνω από μία αποφάσεις σε κάθε χρονική στιγμή. Τέτοια προβλήματα λύνονται με την χρήση αλγορίθμων πολλών πρακτόρων (multi-agent). Σε ένα συνεργατικό πρόβλημα πολλών πρακτόρων υπάρχουν οι πράκτορες όπου το σύνολο τους συμβολίζεται ως $U=\{1,2,\dots,n\}$.

Σε κάθε βήμα t , ο κάθε πράκτορας $u \in U$ επιλέγει μια απόφαση a_t^u με βάση μια πολιτική π^u . Στη συνέχεια οι αποφάσεις όλων των πρακτόρων συνδυάζονται και δημιουργούν την κοινή απόφαση \bar{a}_t . Τέλος όλοι οι πράκτορες δέχονται επιβράβευση $R_t = r(s_t, \bar{a}_t)$ και το περιβάλλον μεταβαίνει στην επόμενη κατάσταση s_{t+1} με βάση την συνάρτηση μετάβασης $p(s_{t+1}|s_t, \bar{a}_t)$ [73].

Μια προφανής λύση για την επίλυση ενός τέτοιου προβλήματος είναι η χρήση ενός μετα-πράκτορα ο οποίος επιλέγει την κοινή απόφαση \bar{a}_t με βάση την πολιτική $\bar{\pi}$ η οποία είναι το διάλυμα που αποτελείται από τις πολιτικές $\pi^u, u \in U$ δηλαδή $\bar{\pi} = (\pi^1, \dots, \pi^n)$. Έτσι ο μετα-πράκτορας μαθαίνει την συνάρτηση $Q(s, \bar{a}) = E^\pi [\sum_{i=t}^T \gamma^i R_i(s_i, \bar{a}_i) | s_t = s, \bar{a}_t = \bar{a}]$. Αυτό έχει σαν αποτέλεσμα τη δυνατότητα χρήσης οποιουδήποτε αλγόριθμου ενισχυτικής μάθησης ενός πράκτορα. Το πρόβλημα είναι ότι σε αυτή τη περίπτωση το σύνολο αποφάσεων \bar{A} προκύπτει από το τη σύνθεση $\prod_{u \in U} A^u$ η οποία αυξάνεται εκθετικά με το πλήθος των πρακτόρων. Επομένως αυτή η προσέγγιση είναι μη πρακτική για εφαρμογές που απαιτούν μεγάλο αριθμό πρακτόρων.

Μια εναλλακτική προσέγγιση η οποία είναι και η πιο συνηθισμένη τεχνική λέγεται ανεξάρτητο Q-learning (IQL). Σε

αυτή τη προσέγγιση ο κάθε πράκτορας $u \in U$ μαθαίνει την δικιά του συνάρτηση Q :

$$Q^u(s, a^u) = E^{\pi^u} \left[\sum_{i=t}^T \gamma^i R_i(s_i, \bar{a}_i) \mid s_t = s, a_t^u = a^u \right]$$

Αυτή η τεχνική ενώ είναι πολύ απλή έχει ένα σημαντικό μειονέκτημα: Το περιβάλλον φαίνεται μη σταθερό για τον κάθε πράκτορα αφού περιέχει τις αποφάσεις και των υπολοίπων πρακτόρων, οι οποίοι ενημερώνουν τις πολιτικές τους ανεξάρτητα από αυτόν. Μπορούμε να αντιμετωπίσουμε αυτό το πρόβλημα με την αποθήκευση κατάλληλης πρόσθετης πληροφορίας στο replay buffer. Όπως ένας πράκτορας μπορεί να χρησιμοποιήσει δεδομένα που έχουν παρθεί όταν η πολιτική του ήταν διαφορετική έτσι μπορεί να χρησιμοποιήσει δεδομένα που έχουν παρθεί από διαφορετικό περιβάλλον[74]. Εφόσον μπορούμε να ξέρουμε την πολιτική του κάθε πράκτορα σε κάθε βήμα της εκπαίδευσης, μπορούμε να ξέρουμε και τον τρόπο με τον οποίο έχει αλλάξει το περιβάλλον. Για να το πετύχουμε αυτό εκφράζουμε την εξίσωση Bellman για έναν πράκτορα με τρόπο με τον οποίο φαίνονται οι πολιτικές των υπόλοιπων πρακτόρων.

$$Q_u^*(s, a_u | \bar{\pi}_{\hat{u}}) = \sum_{\bar{a}_{\hat{u}}} \bar{\pi}_{\hat{u}}(\bar{a}_{\hat{u}} | s) \left[r(s, a_u, \bar{a}_{\hat{u}}) + \gamma \sum_{s'} p(s' | s, a_u, \bar{a}_{\hat{u}}) \max_{a'_u} Q_u^*(s', a'_u) \right]$$

Ο συμβολισμός \hat{u} δηλώνει όλους τους υπόλοιπους πράκτορες $\in U$. Το μη σταθερό μέρος αυτής της εξίσωσης είναι το $\bar{\pi}_{\hat{u}} = \prod_{i \in \hat{u}} \pi_i(a_i | s)$ το οποίο αλλάζει καθώς οι πράκτορες ενημερώνουν τις πολιτικές τους.

Για να γίνει συλλογή πληροφορίας την στιγμή t_c καταγράφουμε το $\bar{\pi}_{\hat{u}}^{t_c}(\bar{a}_{\hat{u}} | s)$ και αποθηκεύουμε την πληροφορία $(s, a_u, r, \pi(\bar{\pi}_{\hat{u}} | s), s')^{t_c}$ στο replay buffer. Την

χρονική στιγμή t_r εκπαιδεύουμε τους πράκτορες με την συνάρτηση απώλειας:

$$L(\theta) = \sum_{i=1}^b \frac{\bar{\pi}_{\hat{u}}^{t_r}(\bar{\alpha}_{\hat{u}}|s)}{\bar{\pi}_{\hat{u}}^{t_i}(\bar{\alpha}_{\hat{u}}|s)} \left[\left(r + \gamma \max_{a'} Q(s', a'; \theta') - Q(s, a; \theta) \right)^2 \right]$$

όπου t_i η στιγμή που συλλέχτηκε το δείγμα i .

Ορίζουμε το επαυξημένο διάνυσμα κατάστασης $\hat{s} = \{s, \bar{\tau}_{\hat{u}}\} \in \hat{S} = S \times T^{n-1}$, το οποίο περιλαμβάνει την αρχική κατάσταση s καθώς και τα ζεύγη παρατήρησης απόφασης των υπόλοιπων πρακτόρων $\bar{\tau}_{\hat{u}}$. Έτσι μπορούμε να ορίσουμε την νέα συνάρτηση επιβράβευσης:

$$\hat{r}(\hat{s}, a) = \sum_{\bar{\alpha}_{\hat{u}}} \bar{\pi}_{\hat{u}}(\bar{\alpha}_{\hat{u}} | \bar{\tau}_{\hat{u}}) p(s' | s, \bar{\alpha}) p_r(\bar{\tau}'_{\hat{u}} | \bar{\tau}_{\hat{u}}, \bar{\alpha}_{\hat{u}}, s').$$

Η οποία μας επιτρέπει να γράψουμε την καινούρια συνάρτηση Bellman ως εξής:

$$\begin{aligned} & Q(\tau, u) \\ &= \sum_{\hat{s}} p(\hat{s} | \tau) \sum_{\bar{\alpha}_{\hat{u}}} \bar{\pi}_{\hat{u}}(\bar{\alpha}_{\hat{u}} | \bar{\tau}_{\hat{u}}) \left[\hat{r}(\hat{s}, a) \right. \\ & \left. + \gamma \sum_{\hat{s}', \tau', \alpha'} p(s' | s, \bar{\alpha}) p_r(\bar{\tau}'_{\hat{u}} | \bar{\tau}_{\hat{u}}, \bar{\alpha}_{\hat{u}}, s') \pi(\alpha', \tau') p_r(\tau' | \tau, \hat{s}, \alpha) Q(\tau', \alpha') \right] \end{aligned}$$

Αυτή η μετατροπή μας παρέχει τη δυνατότητα να αναδείξουμε την εξάρτηση της εξίσωσης του Bellman στο μη σταθερό όρο $\bar{\pi}_{\hat{u}}(\bar{\alpha}_{\hat{u}} | s)$.

Η παραπάνω προσθήκη αυξάνει την ευστάθεια της εκπαίδευσης και οδηγεί σε καλύτερα αποτελέσματα και μεγαλύτερη ταχύτητα σύγκλισης. Παρόλα αυτά η εξάρτηση της

πολιτικής του ενός πράκτορα με τους υπόλοιπους δεν έχει εξαφανιστεί πλήρως καθώς ο όρος $\frac{\pi_{\hat{u}}^{tr}(\bar{a}_{\hat{u}}|s)}{\pi_{\hat{u}}^{ti}(\bar{a}_{\hat{u}}|s)}$ αποτελεί απλά μια προσέγγισή της.

SAC

Μια διαφορετική προσέγγιση από τους παραπάνω αλγόριθμους είναι οι αλγόριθμοι κλίσης πολιτικής (policy gradient). Στους αλγόριθμους κλίσης πολιτικής ο πράκτορας δεν προσπαθεί να μάθει τη συνάρτηση αξίας για κάθε απόφαση της πολιτικής του αλλά την ίδια την πολιτική. Έτσι η έξοδος του νευρωνικού είναι ένας μόνο νευρώνας για κάθε απόφαση σε αντίθεση με τους προηγούμενους αλγόριθμους, όπου η έξοδος ήταν ένα διάνυσμα με την αξία της κάθε απόφασης και άρα υπήρχε ένας νευρώνας για όλες τις δυνατές επιλογές. Επομένως η απόφαση μπορεί πλέον να είναι ένας συνεχής αριθμός, πράγμα που αποτελεί το μεγαλύτερο πλεονέκτημα των μεθόδων policy gradient. Παρόλα αυτά υπάρχουν και ορισμένα μειονεκτήματα όπως η αργή σύγκλιση, η δυσκολία σε μάθηση εκτός πολιτικής και η ανάγκη για πολλά δείγματα[75].

Η συνάρτηση την οποία προσπαθεί να μάθει το νευρωνικό στις μεθόδους policy gradient είναι η εξής:

$$J(\theta) = E \left[\sum_{t=1}^T r_t + 1 \right]$$

Όπου θ τα βάρη του νευρωνικού, r_t η συνάρτηση επιβράβευσης και T ο αριθμός των αποφάσεων σε μια εποχή. Εφόσον ο σκοπός είναι να μεγιστοποιήσουμε την συνάρτηση $J(\theta)$ χρησιμοποιούμε την μέθοδο gradient ascent για την

εύρεση των βέλτιστων βαρών του νευρωνικού: $\theta \leftarrow \theta + \nabla_{\theta} J(\theta)$. Αποδεικνύεται ότι[76]:

$$\nabla_{\theta} J(\theta) = \sum_{t=1}^T \nabla_{\theta} \log \pi_{\theta}(a_t | s_t) G_t$$

Όπου

$$G_t = \sum_{t'=t}^T \gamma^{t'-t} r_{t'}$$

Ήδη μπορούμε να ορίσουμε έναν αλγόριθμο για την εύρεση της βέλτιστης πολιτικής ως εξής:

```

function REINFORCE
  Initialise  $\theta$  arbitrarily
  for each episode  $\{s_1, a_1, r_2, \dots, s_{T-1}, a_{T-1}, r_T\} \sim \pi_{\theta}$  do
    for  $t = 1$  to  $T - 1$  do
       $\theta \leftarrow \theta + \alpha \nabla_{\theta} \log \pi_{\theta}(s_t, a_t) v_t$ 
    end for
  end for
  return  $\theta$ 
end function

```

Εικόνα 7: Reinforce Algorithm

Το πρόβλημα με τον παραπάνω αλγόριθμο είναι η μεγάλη μεταβλητότητα που παρουσιάζει ο λογάριθμος των πιθανοτήτων η οποία προκαλεί με τη σειρά της μεγάλη μεταβλητότητα στην κλίση της συνάρτησης J . Αυτό έχει ως αποτέλεσμα αστάθεια στη μάθηση η οποία μπορεί να οδηγήσει σε αργή ταχύτητα σύγκλισης ή μπορεί το νευρωνικό να μην συγκλίνει ποτέ[77].

Μια λύση στο παραπάνω πρόβλημα είναι οι μέθοδοι κριτή ηθοποιού (actor critic). Μπορούμε να εκφράσουμε το $\nabla_{\theta} J(\theta)$ ως εξής:

$$\nabla_{\theta} J(\theta) = E_{s_0, a_0, \dots, s_t, a_t} [\sum_{t=1}^T \nabla_{\theta} \log \pi_{\theta}(\alpha_t | s_t)] E_{r_{t+1}, s_{t+1}, \dots, r_T, s_T} [G_t].$$

Όμως ο όρος $E_{r_{t+1}, s_{t+1}, \dots, r_T, s_T} [G_t]$ είναι η συνάρτηση $Q(s_t, a_t)$. Επομένως:

$$\nabla_{\theta} J(\theta) = E_{s_0, a_0, \dots, s_t, a_t} \left[\sum_{t=1}^T \nabla_{\theta} \log \pi_{\theta}(\alpha_t | s_t) Q(s_t, a_t) \right]$$

Καταλήγουμε λοιπόν στον αλγόριθμο actor critic ο οποίος αποτελείται από 2 νευρωνικά δίκτυα:

1. Ο κριτής (critic) προσεγγίζει την συνάρτηση αξίας. Αυτή μπορεί να είναι είτε η συνάρτηση αξίας απόφασης (Q function) είτε η συνάρτηση αξίας κατάστασης (V function). Σε πολλές περιπτώσεις ο κριτής μπορεί να υπολογίζει την συνάρτηση πλεονεκτήματος $A(s_t, a_t)$.
2. Ο ηθοποιός (actor) ενημερώνει την κατανομή της πολιτικής στην κατεύθυνση που προτείνει ο κριτής

Algorithm Q Actor Critic

Initialize parameters s, θ, w and learning rates $\alpha_{\theta}, \alpha_w$; sample $a \sim \pi_{\theta}(a|s)$.
for $t = 1 \dots T$: **do**
 Sample reward $r_t \sim R(s, a)$ and next state $s' \sim P(s'|s, a)$
 Then sample the next action $a' \sim \pi_{\theta}(a'|s')$
 Update the policy parameters: $\theta \leftarrow \theta + \alpha_{\theta} Q_w(s, a) \nabla_{\theta} \log \pi_{\theta}(a|s)$; Compute the correction (TD error) for action-value at time t :
 $\delta_t = r_t + \gamma Q_w(s', a') - Q_w(s, a)$
 and use it to update the parameters of Q function:
 $w \leftarrow w + \alpha_w \delta_t \nabla_w Q_w(s, a)$
 Move to $a \leftarrow a'$ and $s \leftarrow s'$
end for

Εικόνα 8: Q Actor Critic Algorithm

Ο αλγόριθμος Actor critic είναι σταθερός και τελικά συγκλίνει σε μια τοπικά ή ολικά βέλτιστη λύση. Παρόλα αυτά εξακολουθεί να πάσχει από βραδεία εκπαίδευση. Αυτό προκύπτει επειδή μετά από κάθε ενημέρωση του νευρωνικού αλλάζει η πολιτική και επομένως τα δεδομένα που έχουν συλλεχθεί δεν φέρουν πλέον χρήσιμη πληροφορία και άρα χρειάζεται νέα συλλογή δεδομένων[78]. Για αυτό το λόγο έχουν αναπτυχθεί παραλλαγές του αλγορίθμου με σκοπό την καλύτερη αξιοποίηση δεδομένων ακόμα και μετά την κάθε ενημέρωση. Μια τέτοια παραλλαγή είναι ο αλγόριθμος Soft Actor Critic (SAC).

Ο αλγόριθμος SAC δίνει παραπάνω επιβράβευση σε πολιτικές οι οποίες εμφανίζουν μεγαλύτερη εντροπία. Η εντροπία (entropy) ορίζεται ως η αβεβαιότητα που εμφανίζει μια τυχαία μεταβλητή P ως εξής:

$$H(P) = E_{x \in P}[-\log P(x)].$$

Με βάση αυτόν τον ορισμό, η βέλτιστη πολιτική που προσπαθεί να υπολογίσει ένα σύστημα ενισχυτικής μάθησης είναι

$$\pi^* = \operatorname{argmax}_{\pi} \left[\sum_{t=1}^T \gamma^t ((R(s_t, a_t, s_{t+1}) + aH(\pi(\cdot | s_t))) \right]$$

Όπου a μια υπερπαραμέτρος του συστήματος που λέγεται συντελεστής αντιστάθμισης (trade-off coefficient). Οι νέοι ορισμοί των συναρτήσεων V και Q είναι οι εξής:

$$V^{\pi}(s) = \sum_{t=1}^T \gamma^t (R(s_t, a_t, s_{t+1}) + a H(\pi(\cdot | s_t))) | s_0 = s$$

$$Q^{\pi}(s, a)$$

$$= E(\sum_{t=1}^T \gamma^t R(s_t, a_t, s_{t+1}) + a \sum_{t=1}^T \gamma^t H(\pi(\cdot | s_t)) | s_0 = s, a_0 = a)$$

Η νέα εξίσωση του Bellman για την εύρεση της βέλτιστης πολιτικής με βάση τον ορισμό της εντροπίας είναι:

$$\begin{aligned} Q^\pi(s, a) &= E_{s' \in P} \left[R(s, a, s') + \gamma \left(Q^\pi(s', a') + aH(\pi(\cdot | s')) \right) \right] \\ &= E_{s' \in P} [R(s, a, s') + \gamma(Q^\pi(s', a') - a \log \pi(a' | s'))] \\ &= E_{s' \in P} [R(s, a, s') + \gamma V^\pi(s')] \end{aligned}$$

Στην πράξη μπορούμε να προσεγγίσουμε αυτή την εξίσωση με δείγματα αφού πρόκειται για μέση τιμή ως εξής:

$$Q^\pi(s, a) \approx r + \gamma(Q^\pi(s', a') - a \log \pi(a' | s'))$$

Μπορούμε πλέον να εκφράσουμε τη συνάρτηση απώλειας του νευρωνικού του κριτή ως εξής:

$$Loss(\theta) = E[(Q_\theta(s, a) - y(r, s', d))^2]$$

Όπου:

$$y(r, s', d) = r + \gamma(1 - d) \left(\min_{j=1,2} Q_{\theta_{targ,j}}(s', a') - a \log \pi_\theta(a', s') \right)$$

Ο αλγόριθμος SAC επομένως μπορεί να γραφτεί ως:

Algorithm Soft Actor-Critic

1: Input: initial policy parameters θ , Q-function parameters ϕ_1, ϕ_2 , empty replay buffer \mathcal{D}
2: Set target parameters equal to main parameters $\phi_{\text{target},1} \leftarrow \phi_1, \phi_{\text{target},2} \leftarrow \phi_2$
3: **repeat**
4: Observe state s and select action $a \sim \pi_\theta(\cdot|s)$
5: Execute a in the environment
6: Observe next state s' , reward r , and done signal d to indicate whether s' is terminal
7: Store (s, a, r, s', d) in replay buffer \mathcal{D}
8: If s' is terminal, reset environment state.
9: **if** it's time to update **then**
10: **for** j in range(however many updates) **do**
11: Randomly sample a batch of transitions, $B = \{(s, a, r, s', d)\}$ from \mathcal{D}
12: Compute targets for the Q functions:

$$y(r, s', d) = r + \gamma(1 - d) \left(\min_{i=1,2} Q_{\phi_{\text{target},i}}(s', \tilde{a}') - \alpha \log \pi_\theta(\tilde{a}'|s') \right), \quad \tilde{a}' \sim \pi_\theta(\cdot|s')$$

13: Update Q-functions by one step of gradient descent using

$$\nabla_{\phi_i} \frac{1}{|B|} \sum_{(s,a,r,s',d) \in B} (Q_{\phi_i}(s, a) - y(r, s', d))^2 \quad \text{for } i = 1, 2.$$

14: Update policy by one step of gradient ascent using

$$\nabla_\theta \frac{1}{|B|} \sum_{s \in B} \left(\min_{i=1,2} Q_{\phi_i}(s, \tilde{a}_\theta(s)) - \alpha \log \pi_\theta(\tilde{a}_\theta(s)|s) \right),$$

where $\tilde{a}_\theta(s)$ is a sample from $\pi_\theta(\cdot|s)$ which is differentiable wrt θ via the reparametrization trick.
15: Update target networks with

$$\phi_{\text{target},i} \leftarrow \rho \phi_{\text{target},i} + (1 - \rho) \phi_i \quad \text{for } i = 1, 2$$

16: **end for**
17: **end if**
18: **until** convergence

Εικόνα 9: Soft Actor Critic Algorithm[79]

Κεφάλαιο 7, Φορμαλισμός του Προβλήματος

Ορισμός του προβλήματος

Σε αυτή την εργασία θεωρούμε ένα πάρκινγκ ηλεκτρικών αυτοκινήτων ως ένα σύστημα V2G. Συγκεκριμένα μοντελοποιούμε το πρόβλημα από την πλευρά του συσσωρευτή και δοκιμάζουμε διαφορετικούς αλγορίθμους ενισχυτικής μάθησης με σκοπό την σύγκρισή τους ως προς το κέρδος που μπορούν να επιφέρουν. Θεωρούμε ένα πάρκινγκ με χωρητικότητα A ηλεκτρικών αυτοκινήτων το οποίο διαθέτει ικανότητα φόρτισής τους. Το κάθε αυτοκίνητο όταν έρχεται δηλώνει τον χρόνο της διαμονής του στο παρκινγκ καθώς και το ποσοστό της μπαταρίας που πρέπει να έχει κατά την αποχώρισή του. Το κάθε αυτοκίνητο i λοιπόν μπορεί να αναπαρασταθεί με μια τούπλα:

$$\langle b_i, d_i, p_i^{\min}, p_i^{\max}, p_{i,t}^{\min}, p_{i,t}^{\max}, p_{total}, SOC_{current}, SOC_{desired} \rangle$$

Όπου:

- b_i : Η χρονική στιγμή που ο οδηγός θέλει να αποχωρήσει. Ο περιορισμός αυτός δεν είναι αυστηρός υπάρχει δηλαδή η δυνατότητα αποχώρησης μετά τη στιγμή b_i στην οποία περίπτωση όμως το πάρκινγκ θα πρέπει να αποζημιώσει τον οδηγό για την οποιαδήποτε καθυστέρηση.
- d_i : Η χρονική στιγμή που ο οδηγός είναι απαραίτητο να αποχωρήσει. Πρόκειται για έναν αυστηρό περιορισμό τον οποίον ο αλγόριθμος φόρτισης είναι υποχρεωμένος να τηρήσει. Ισχύει προφανώς $b_i \leq d_i$.
- p_i^{\max} : Η μέγιστη τιμή που μπορεί να πάρει η ενέργεια η οποία παρέχεται στο αυτοκίνητο σε ένα χρονικό διάστημα. Πρόκειται για έναν περιορισμό που εισάγεται από τα χαρακτηριστικά της μπαταρίας του αυτοκινήτου.

- p_i^{min} : Η ελάχιστη τιμή που μπορεί να πάρει η ενέργεια η οποία παρέχεται στο αυτοκίνητο. Εφόσον εφαρμόζουμε τεχνικές V2G, η τιμή της μπορεί να είναι και αρνητική και άρα συμβολίζει την μέγιστη ενέργεια που μπορεί να δώσει το αυτοκίνητο στο δίκτυο σε ένα χρονικό διάστημα.
- $p_{i,t}^{min}$: Η ελάχιστη ενέργεια που μπορεί να δεχτεί (ή η μέγιστη που μπορεί να δώσει στο δίκτυο) το αυτοκίνητο την στιγμή t ώστε να καλύπτονται οι περιορισμοί. Το σύστημα περιορισμών με το οποίο προκύπτει είναι το εξής:
 - $p_{i,t}^{min} \leq p_i^{min} \forall t$
 - $SOC_{current} + PtoSOC(p_{i,t}^{min}) \geq 0$
 - Έστω P_{next}^{max} η μέγιστη ενέργεια που μπορεί να δοθεί στο αυτοκίνητο στο χρόνο μέχρι την αυστηρή αποχώρησή του. Πρέπει: $SOC_{current} + PtoSOC(p_{i,t}^{min}) + PtoSOC(p_{next}^{max}) \geq SOC_{desired}$
- $p_{i,t}^{max}$: Η μέγιστη ενέργεια που μπορεί να δεχτεί το αυτοκίνητο τη στιγμή t . Ο μόνος περιορισμός είναι αυτός που προκαλείται από τη μέγιστη χωρητικότητα, δηλαδή $SOC_{current} + PtoSOC(p_{t,t}^{max}) \leq 1$
- p_{total} : Η μέγιστη ενέργεια που μπορεί να αποθηκεύσει το αυτοκίνητο, δηλαδή η χωρητικότητα της μπαταρίας του.
- $SOC_{current}$: Το επίπεδο φόρτισης του αυτοκινήτου.
- $SOC_{desired}$: Το επίπεδο φόρτισης που πρέπει να έχει το αυτοκίνητο τη στιγμή της αποχώρησής του.

Στην περίπτωση που παραβιαστεί ο ελαστικός χρονικός περιορισμός ενός αυτοκινήτου η ποινή που δέχεται το πάρκινγκ είναι $U(t) = \frac{\delta^{t-b_i} SOC_{current}}{SOC_{desired}}$.

Σε κάθε χρονική στιγμή ο συσσωρευτής καλείται να αποφασίσει την ενέργεια που θα δώσει ή θα πάρει από το κάθε αυτοκίνητο. Χρειάζεται δηλαδή συνολικά $P = \sum_{i=1}^A p_i \text{KW}$ ενέργειας. Η τιμή του P μπορεί να είναι και αρνητική στην περίπτωση που η συνολική ενέργεια που δέχεται από τα αυτοκίνητα είναι περισσότερη από αυτή που παρέχει. Η ενέργεια πωλείται ή αγοράζεται από το δίκτυο με κόστος l_t το οποίο διαφοροποιείται κατά τη διάρκεια του 24ώρου.

Ο συσσωρευτής παίρνει υπόψη του και τη φθορά που δέχεται η μπαταρία του κάθε αυτοκινήτου από τις επιλογές του. Η συνάρτηση με την οποία υπολογίζεται το κόστος της φθοράς είναι: $C_{deg} = C_T + C_{c-rate} + C_{SOC}$ όπου:

- C_T : Η φθορά που προκαλείται λόγω της θερμοκρασίας. Σε αυτό το μοντέλο θεωρούμε ότι η θερμοκρασία του αυτοκινήτου είναι ανεξάρτητη των αποφάσεων του συσσωρευτή και για αυτό το λόγο δεν λαμβάνεται υπόψη. Θεωρούμε δηλαδή ότι $C_T = 0$.
- C_{c-rate} : Η φθορά που προκαλείται λόγω του υψηλού ρυθμού φόρτισης. Αυτό ορίζεται ως: $C_{c-rate} = \frac{\delta_{crate}}{c^{rate}}$ όπου c^{rate} σταθερά η οποία προκύπτει από τον Πίνακα 2: απαραίτητοι κύκλοι φόρτισης μέχρι το τέλος ζωής μιας ηλεκτρικής μπαταρίας για διαφορετικούς ρυθμούς φόρτισης. Αν δηλαδή για μια τιμή του ρυθμού φόρτισης χρειάζονται c κύκλοι για να φτάσει η μπαταρία στο τέλος της ζωής της το κόστος που υπολογίζει ο συσσωρευτής είναι ανάλογο του $\frac{1}{c}$. Για τιμές ανάμεσα στις ενδεικτικές θεωρούμε γραμμική μεταβολή. Στην περίπτωση που το αυτοκίνητο αποφορτίζεται στο δίκτυο θεωρούμε ότι $C_{c-rate} = 0$. Ο όρος δ_{crate} είναι σταθερά και επιλέγεται με βάση την προτεραιότητα που

ο διαχειριστής του συσσωρευτή θέλει να δοθεί στον συγκεκριμένο όρο.

C-rate	0.5C	1C	2C	3C	4C	5C
Cycles to reach End of Life	4000	2875	2000	1250	875	750

Πίνακας 2: απαραίτητοι κύκλοι φόρτισης μέχρι το τέλος ζωής μιας ηλεκτρικής μπαταρίας για διαφορετικούς ρυθμούς φόρτισης

- C_{SOC} : Η φθορά που προκαλείται λόγω του επιπέδου φόρτισης του αυτοκινήτου. Ορίζεται ως $C_{SOC} = \frac{SOC_{cost}}{900 \cdot 24 \cdot 60 \cdot 60 \cdot 0.2} \delta_{SOC}$. Η σταθερά SOC_{cost} προκύπτει από τον Πίνακας 3: Απώλεια χωρητικότητας ηλεκτρικής μπαταρίας σε διαφορετικά επίπεδα φόρτισης. Πρόκειται για τα αποτελέσματα ενός πειράματος στο οποίο μετρήθηκε η απώλεια χωρητικότητας που προκλήθηκε σε αυτοκίνητα σε διάστημα 900 ημερών[80]. Η σταθερά $900 \cdot 24 \cdot 60 \cdot 60$ είναι ο αριθμός των δευτερολέπτων σε 900 μέρες ενώ η σταθερά 0.2 είναι η μέγιστη απώλεια χωρητικότητας που μπορεί να δεχτεί μια μπαταρία μέχρι να θεωρείται ότι έφτασε στο τέλος της ζωής της. Η σταθερά δ_{SOC} επιλέγεται από το διαχειριστή του συσσωρευτή παρόμοια με τη δ_{C-rate} .

Column1	Column2	Column3	Column4
State of Charge	0%	50%	100%
Capacity Loss	2%	5%	11%

Πίνακας 3: Απώλεια χωρητικότητας ηλεκτρικής μπαταρίας σε διαφορετικά επίπεδα φόρτισης

Σχεδίαση Μοντέλου

Για να επιλυθεί το πρόβλημα λήψης αποφάσεων του συσσωρευτή με ενισχυτική μάθηση πρέπει να μοντελοποιηθεί ως μια διαδικασία απόφασης Markov.

Ως διάνυσμα κατάστασης του μοντέλου ορίζουμε την τούπλα $\langle t, \bar{\Omega}, l_t \rangle$ όπου t η χρονική στιγμή της απόφασης, l_t η τιμή του ρεύματος σε kw την χρονική στιγμή t και $\bar{\Omega}$ ένα σύνολο από τούπλες αυτοκινήτων μήκους A . Οι μεταβάσεις από μια κατάσταση στην επόμενη είναι οι εξής:

1. $t \rightarrow t + 1$
2. $l_t \rightarrow l_{t+1}$
3. $\bar{\Omega} \rightarrow \bar{\Omega}'$ όπου $\bar{\Omega}'$ το σύνολο αυτοκινήτων που προκύπτει από τις αποφάσεις του συσσωρευτή, τις αποχωρίσεις και τις καινούριες αφίξεις με βάση κάποια κατανομή αφίξεων.

Όταν τα αυτοκίνητα που βρίσκονται στο πάρκινγκ είναι λιγότερα από A το σύνολο Ω συμπληρώνεται με μηδενικά. Η αρχική κατάσταση είναι $\langle 0, 0, \dots, 0, l_0 \rangle$.

Η συνάρτηση επιβράβευσης ορίζεται ως:

$$r(t) = - \sum_{i \in A} C_{deg}(SOC_{current,i}, C_{rate,i}) - \sum_{i \in A} U(t, SOC_{current,i}) - \sum_{i \in A} l_t * p_{i,t}$$

Όπου p_i η ενέργεια που παρέχει ο συσσωρευτής στο αυτοκίνητο i τη στιγμή t .

Το παραπάνω μοντέλο μπορεί να χρησιμοποιηθεί ως περιβάλλον σε ένα σύστημα ενισχυτικής μάθησης με πολλούς πράκτορες. Για την περίπτωση του ενός πράκτορα όμως θα πρέπει να γίνουν ορισμένες αλλαγές. Ορίζουμε ως νέο state την τούπλα $\langle t, i, \Omega_i, \bar{\Omega}, l_t \rangle$ προσθέτουμε δηλαδή τα i, Ω_i . Το i είναι το πλήθος των αυτοκινήτων για τα οποία πρέπει να παρθεί απόφαση τη χρονική στιγμή t . Το Ω_i είναι το

αυτοκίνητο για το οποίο παίρνει απόφαση ο συσσωρευτής και το $\bar{\Omega}$ είναι όλα τα υπόλοιπα αυτοκίνητα. Οι μεταβάσεις είναι:

1. $t \rightarrow t + 1$
2. $l_t \rightarrow l_{t+1}$
3. $\bar{\Omega} \rightarrow \bar{\Omega}'$
4. $i \rightarrow |\bar{\Omega}'|$
5. $\Omega_i \rightarrow \Omega_i'$ όπου Ω_i' το πρώτο αυτοκίνητο του καινούριου συνόλου.

Όταν $i = 0$ και:

1. $t \rightarrow t$
2. $l_t \rightarrow l_t$
3. $\bar{\Omega} \rightarrow \bar{\Omega}'$
4. $i \rightarrow i - 1$
5. $\Omega_i \rightarrow \Omega_i'$ όπου Ω_i' το επόμενο αυτοκίνητο του συνόλου.

Όταν $i > 0$. Η αρχική κατάσταση είναι όμοια με το μοντέλο πολλών πρακτόρων με ένα παραπάνω μηδενικό στη θέση του i . Η συνάρτηση επιβράβευσης είναι

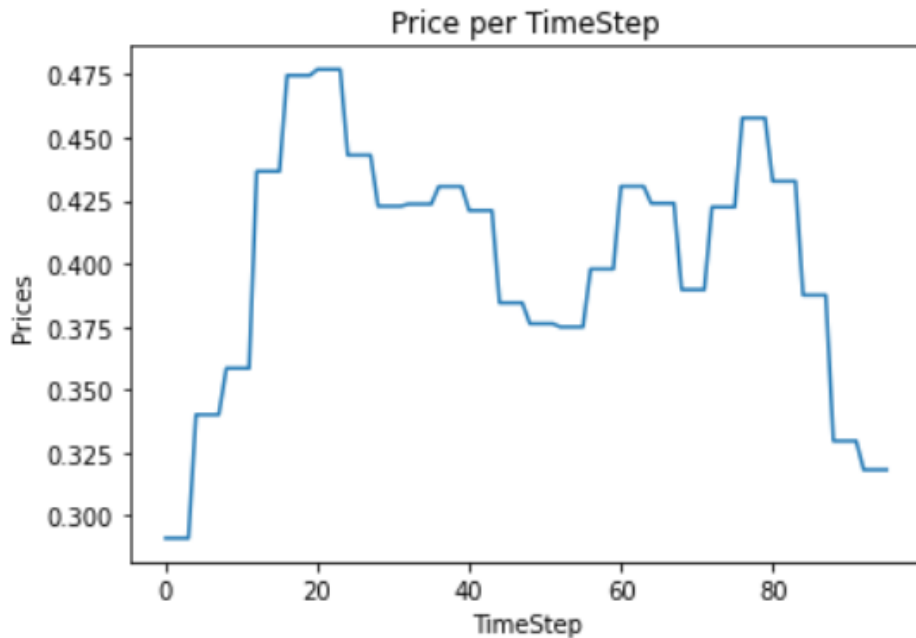
$$r(t, i) = \begin{cases} -C_{deg}(SOC_{current,i}, C_{rate,i}) - U(t, SOC_{current,i}) - l_t * p_i, & i > 0 \\ 0, & i = 0 \end{cases}$$

Και στα δύο παραπάνω μοντέλα θεωρούμε χρονικό διάστημα μιας εποχής τη μία μέρα (24 ώρες) και χρονικό παράθυρο μιας απόφασης τα 15 λεπτά.

Όσον αφορά το πλήθος των αυτοκινήτων A έγινε δοκιμή σε 3 διαφορετικές τιμές του: 20, 50 και 70. Αυτό επιλέχθηκε ώστε οι αλγόριθμοι επίλυσης του προβλήματος να δοκιμαστούν και όσο αφορά τη κλιμακωσιμότητά τους.

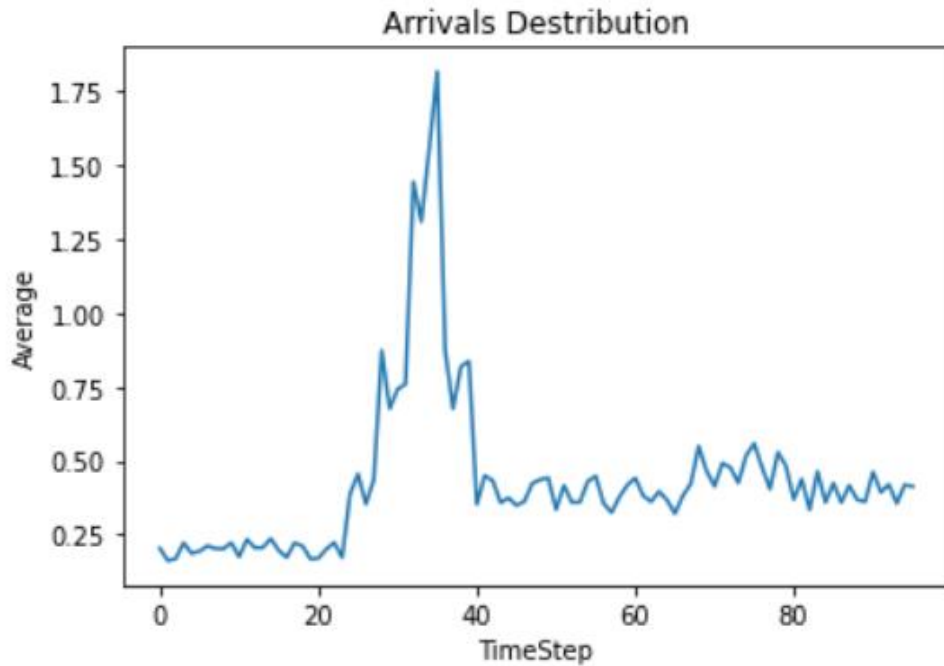
Παραγωγή δεδομένων

Το κόστος της ηλεκτρικής ενέργειας τη στιγμή t , l_t , παράχθηκε από τυχαία κανονική κατανομή με μέση τιμή $\mu(l_t)$ και διασπορά $\sigma(l_t) = 0.03$. Η αρχική τιμή του $\mu(l_0)$ τέθηκε ίση με $0.4\$/kW$. Για τις υπόλοιπες χρονικές στιγμές θεωρήσαμε ότι η μέση τιμή $\mu(l_t)$ ακολουθεί αλυσίδα Markov όπου $\mu(l_t) = l_{t-1} + 0.02$. Παρακάτω βλέπουμε ένα παράδειγμα κατανομής του ημερήσιου κόστους ενέργειας. Παρατηρούμε ότι η κατανομή που παράγεται με αυτή τη διαδικασία μοιάζει αρκετά με αυτή της *Εικόνα*.



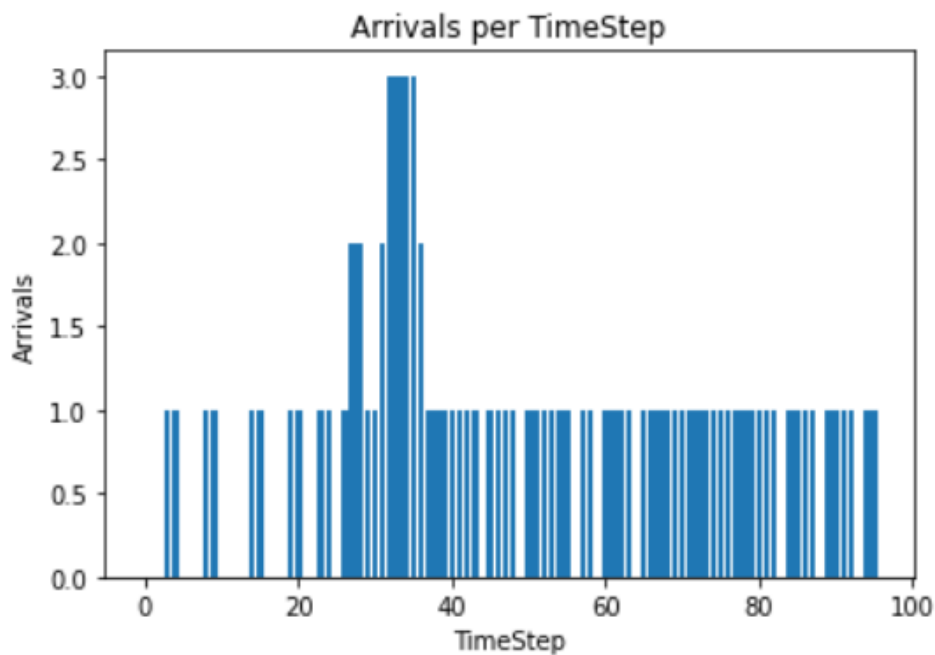
Εικόνα 10: Κόστος ηλεκτρικού ρεύματος ανά χρονική στιγμή

Οι αφίξεις των αυτοκινήτων έγιναν με τυχαία κατανομή η μέση τιμή της οποίας διαφέρει σε κάθε χρονική στιγμή. Τις πρωινές ώρες η πιθανότητα των αφίξεων είναι μικρή, το μεσημέρι είναι πολύ υψηλή και το απόγευμα είναι πάλι μικρότερη.



Εικόνα 11: Κατανομή αφίξεων ηλεκτρικών αυτοκινήτων

Στην εικόνα 12 φαίνεται ο αριθμός των αφίξεων σε κάθε χρονικό παράθυρο μέσα σε ένα εικοσιτετράωρο.



Εικόνα 12: Αφίξεις ηλεκτρικών αυτοκινήτων ανά χρονική στιγμή

Τα χαρακτηριστικά του αυτοκινήτου παράγονται ως εξής:

- Η ώρα αποχώρησης b είναι τυχαία μεταβλητή ομοιόμορφη στο διάστημα (στιγμή άφιξης, τέλος ημέρας)].
- Η ώρα αυστηρής αποχώρησης d είναι ομοιόμορφη τυχαία μεταβλητή στο διάστημα $[b, \text{τέλος ημέρας}]$.
- Το επίπεδο φόρτισης $SOC_{current}$ είναι τυχαία μεταβλητή στο διάστημα $[0,1]$.
- Το επιθυμητό επίπεδο φόρτισης $SOC_{desired}$ είναι τυχαία μεταβλητή στο διάστημα $(SOC_{current}, 1]$ με τον περιορισμό του χρόνου. Θεωρούμε δηλαδή ότι ο χρόνος διαμονής του αυτοκινήτου είναι αρκετός για την φόρτισή του στο επιθυμητό επίπεδο.
- Η συνολική ενέργεια σε kW που μπορεί να αποθηκεύσει η μπαταρία του αυτοκινήτου είναι ανάμεσα σε 17.6 και 100kW.
- Η τιμή του μέγιστου ρυθμού φόρτισης και αποφόρτισης του αυτοκινήτου είναι ο ρυθμός φόρτισης c-rate πολλαπλασιασμένος με μια τυχαία μεταβλητή από το 1 έως το 5.

Αλγόριθμοι Απόφασης

Βασικός αλγόριθμος

Ο πρώτος αλγόριθμος που δοκιμάστηκε για την βέλτιστη επίλυση του προβλήματος απόφασης του συσσωρευτή είναι ο βασικός αλγόριθμος φόρτισης που δεν κάνει χρήση καμίας τεχνικής μηχανικής μάθησης. Κάθε αυτοκίνητο φορτίζεται με σταθερό ρυθμό από την άφιξή του στο πάρκινγκ μέχρι την αποχώρισή του από αυτό. Όπως είναι προφανές ο αλγόριθμος αυτός δεν εμφανίζει ιδιαίτερο ενδιαφέρον αλλά χρησιμοποιήθηκε μόνο ως μέτρο σύγκρισης των αλγορίθμων

τεχνητής νοημοσύνης με κάτι σταθερό που δεν εμφανίζει τυχειότητα ή αστάθεια.

Με βάση την λύση αυτού του αλγορίθμου τέθηκαν οι τιμές των συντελεστών των συναρτήσεων κόστους, με σκοπό η φθορά που προκαλείται στην μπαταρία να είναι περίπου 2 φορές μεγαλύτερη από το κόστος του ρεύματος. Η επιλογή αυτή έγινε αυθαίρετα για να τεθεί προτεραιότητα στις επιπτώσεις των αποφάσεων που λαμβάνει ο κάθε αλγόριθμος.

DQN

Ο δεύτερος αλγόριθμος που δοκιμάστηκε ήταν ο αλγόριθμος DQN. Το νευρωνικό Q που επιλέχθηκε αποτελείται από το επίπεδο εισόδου, 2 κρυφά επίπεδα και το επίπεδο εξόδου. Ανάμεσα σε κάθε κρυφό επίπεδο προστέθηκε ένα επίπεδο κανονικοποίησης (Batch Normalization) και ένα επίπεδο εγκατάλειψης (Dropout) με παράμετρο 0.3. Το επίπεδο εξόδου είχε 21 νευρώνες, δηλαδή 21 πιθανές αποφάσεις. Το πλήθος νευρώνων του κάθε επιπέδου φαίνονται στον πίνακα 4.

A	Activation Layer	Hidden Layer 1	Hidden Layer 2	Output Layer
20	183	1000	500	21
50	453	1500	800	21
70	633	2000	1000	21

Πίνακας 4: Πλήθος νευρώνων κάθε επιπέδου του νευρωνικού δικτύου του αλγορίθμου DQN

Η απόφαση του νευρωνικού ήταν ένας αριθμός a_i για το κάθε αυτοκίνητο στο διάστημα $[0,1]$. Αυτός ο αριθμός κανονικοποιούνταν στο διάστημα $[p_{i,t}^{min}, p_{i,t}^{max}]$ και έτσι προέκυπτε το ρεύμα σε kW $p_{i,t}$. Εφόσον ο αλγόριθμος DQN έχει συνεχές σύνολο αποφάσεων και οι νευρώνες απόφασης

είναι 21, ο a_i πρέπει να είναι κάποιο πολλαπλάσιο του 0.05 ανήκει δηλαδή στο $[0, 0.05, 0.1, \dots, 1]$.

Ως συνάρτηση βελτιστοποίησης του πράκτορα επιλέχθηκε η συνάρτηση Adam, ως συνάρτηση απώλειας επιλέχθηκε η συνάρτηση τετραγώνου (Squared Loss). Το μέγεθος παρτίδας (Batch size) του replay buffer τέθηκε στο 64. Το μέγεθος του replay buffer τέθηκε στα 100.000 δείγματα.

IQI

Ο επόμενος αλγόριθμος που δοκιμάστηκε ήταν ο αλγόριθμος IQI. Σε αυτή τη περίπτωση το πλήθος των πρακτόρων του αλγορίθμου ήταν όσο το πλήθος των αυτοκινήτων A. Το νευρωνικό δίκτυο Q του κάθε πράκτορα αποτελούταν από 2 κρυφά επίπεδα. Το Επίπεδο εξόδου του κάθε νευρωνικού είχε 21 νευρώνες.

A	Activation Layer	Hidden Layer 1	Hidden Layer 2	Output Layer
20	183	1000	1000	21
50	453	1500	1500	21
70	633	2000	2000	21

Πίνακας 5: Πλήθος νευρώνων κάθε επιπέδου του νευρωνικού δικτύου του αλγορίθμου IQI

Η απόφαση του κάθε πράκτορα ήταν ένας αριθμός από το 0 στο 1 με βήμα 0.05 όπως του DQN. Η διαφορά είναι ότι πλέον οι αποφάσεις για τη κάθε θέση στάθμευσης παίρνονται παράλληλα με τον κάθε πράκτορα να αποφασίζει για μια θέση.

Ως συνάρτηση βελτιστοποίησης επιλέχθηκε η συνάρτηση Adam. Το μέγεθος παρτίδας τέθηκε στο 64. Το μέγεθος του replay buffer τέθηκε στο 1.000.000 δείγματα.

SAC

Ο τελευταίος αλγόριθμος που δοκιμάστηκε ήταν ο αλγόριθμος SAC. Για τα νευρωνικά δίκτυα και του κριτή και του ηθοποιού χρησιμοποιήθηκε από ένα νευρωνικό με 2 κρυφά επίπεδα, με πλήθος νευρώνων ίδιο με αυτό του DQN. Το επίπεδο εξόδου είχε έναν νευρώνα δηλαδή την απόφαση του κάθε δικτύου.

Η απόφαση του πράκτορα ήταν ένας συνεχής αριθμός από το 0 μέχρι το 1, Ο οποίος κανονικοποιήθηκε για κάθε αυτοκίνητο στο διάστημα $[p_{i,t}^{min}, p_{i,t}^{max}]$ για τον υπολογισμό του ρεύματος $p_{i,t}$.

Ως συνάρτηση βελτιστοποίησης κάθε δικτύου επιλέχθηκε η συνάρτηση Adam. Η ίδια συνάρτηση επιλέχθηκε για την βελτιστοποίηση της παραμέτρου α του αλγορίθμου SAC. Η τιμή της περιόδου μάθησης του κριτή τέθηκε στη μία εποχή. Ως συνάρτηση απώλειας τέθηκε η συνάρτηση διαφοράς τετραγώνων.

Κεφάλαιο 8, Πειράματα και Αποτελέσματα

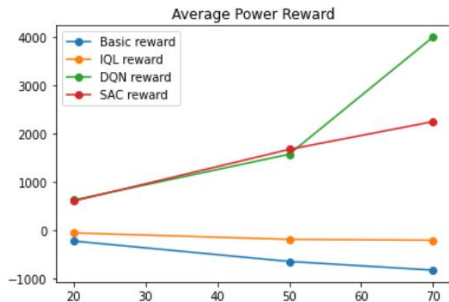
Παράθεση των πειραμάτων

Μετά την επιτυχημένη εκπαίδευση των αλγορίθμων μηχανικής μάθησης, και αφού έχουν συγκλίνει σε κάποια ικανοποιητική λύση, πραγματοποιήθηκαν 3 πειράματα για την αξιολόγησή τους, καθώς και την σύγκρισή τους με τον βασικό αλγόριθμο. Στο κάθε πείραμα παράχθηκαν 100 διαφορετικές περιπτώσεις του προβλήματος και οι αλγόριθμοι κλήθηκαν να βρουν το βέλτιστο σύνολο αποφάσεων για την επίλυσή τους. Η σύγκριση των αλγορίθμων έγινε πρώτα με βάση την μέση συνολική επιβράβευση που αποτίμησαν οι αλγόριθμοι, ύστερα με βάση την μέση επιβράβευση που αποτίμησαν μόνο με κατάλληλη αγορά και πώληση ρεύματος, και τέλος με βάση μόνο τη μέση φθορά που προκάλεσαν οι αποφάσεις τους στις μπαταρίες των ηλεκτρικών αυτοκινήτων.

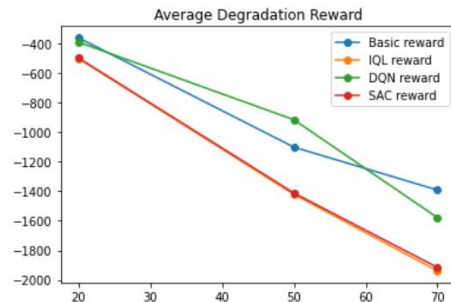
Στο πρώτο πείραμα η κατανομή άφιξης των αυτοκινήτων ήταν ίδια με αυτή στην οποία εκπαιδεύτηκαν τα νευρωνικά δίκτυα των πρακτόρων. Στο δεύτερο πείραμα έγινε δοκιμή σε κατανομή παρόμοιας μορφής αλλά με σημαντικά αυξημένο ρυθμό αφίξεων. Στο τελευταίο πείραμα έγινε δοκιμή σε κατανομή με μικρότερο ρυθμό αφίξεων αυτοκινήτων. Τέλος έγινε επανάληψη όλων των πειραμάτων και για τα 3 μεγέθη του πάρκινγκ ηλεκτρικών αυτοκινήτων ώστε να γίνει κατάλληλη σύγκριση της κλιμακωσιμότητας των αλγορίθμων.

Πείραμα Ίσης κατανομής

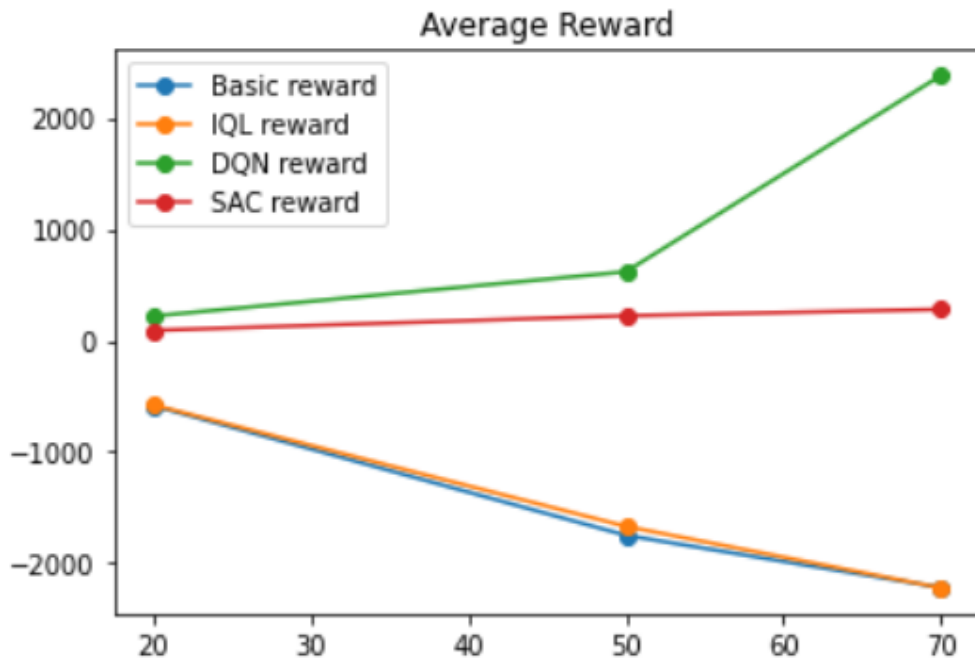
Παρακάτω φαίνονται τα αποτελέσματα του πειράματος για διαφορετικά μεγέθη παρκινγκ.



Εικόνα 13: Επιβράβευση λόγω αγοράς ηλεκτρικού ρεύματος σε πείραμα ίσης κατανομής



Εικόνα 14: Επιβράβευση λόγω φθοράς της μπαταρίας σε πείραμα ίσης κατανομής



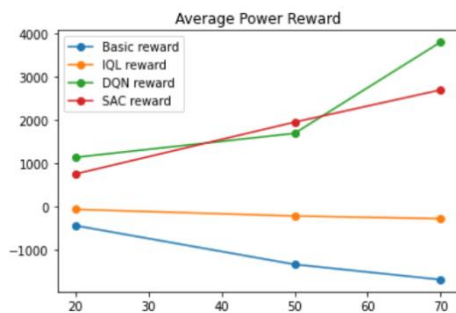
Εικόνα 15: Συνολική επιβράβευση σε πείραμα ίσης κατανομής

Το συνολικό αποτέλεσμα δείχνει ότι η καλύτερη λύση υπολογίστηκε από τον αλγόριθμο DQN. Ο αλγόριθμος SAC έφτασε και αυτός σε αρκετά ικανοποιητικό αποτέλεσμα ενώ ο

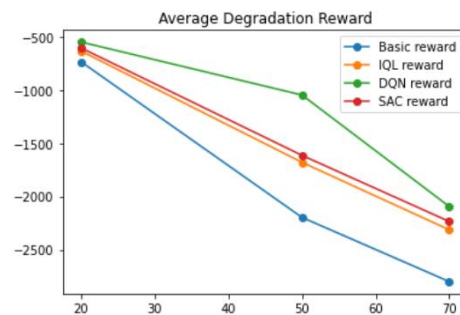
αλγόριθμος IQL δεν κατάφερε να ξεπεράσει τον βασικό αλγόριθμο. Όσο αφορά την φθορά της μπαταρίας ο αλγόριθμος DQN έφτασε σε αποτέλεσμα που -αν και ελαφρώς χειρότερο- δεν διαφέρει σημαντικά από αυτό του βασικού αλγορίθμου. Οι αλγόριθμοι SAC και IQL προκαλούν αρκετά περισσότερη φθορά από τον βασικό αλγόριθμο. Παρόλα αυτά το κέρδος που επέφερε ο SAC είναι αρκετό για να αντισταθμίσει την φθορά που προκάλεσαν στις μπαταρίες των αυτοκινήτων οι αποφάσεις του.

Πείραμα κατανομής Περισσότερων Αυτοκινήτων

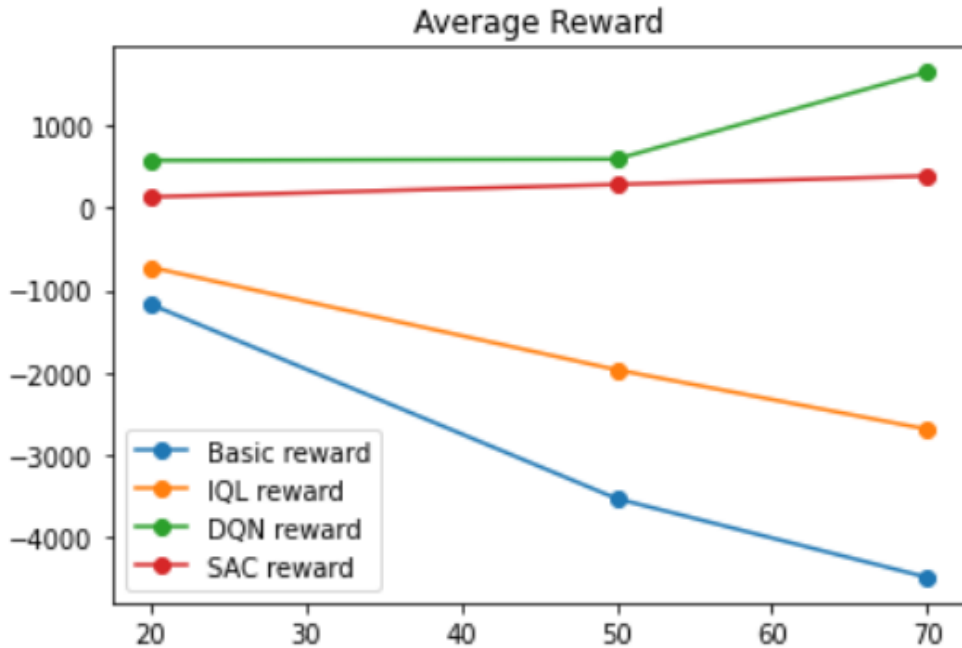
Το δεύτερο πείραμα πραγματοποιήθηκε με κατανομή αυτοκινήτων η οποία είχε μεγαλύτερο ρυθμό αφίξεων. Συγκεκριμένα η μορφή της κατανομής ήταν παρόμοια με την προηγούμενη αλλά οι αφίξεις σε κάθε χρονική στιγμή ήταν περίπου διπλάσιες. Παρακάτω φαίνονται τα αποτελέσματα αυτού του πειράματος.



Εικόνα 16: Επιβράβευση λόγω αγοράς ηλεκτρικού ρεύματος σε πείραμα μεγαλύτερης κατανομής



Εικόνα 17: Επιβράβευση λόγω φθοράς της μπαταρίας σε πείραμα ίσης κατανομής

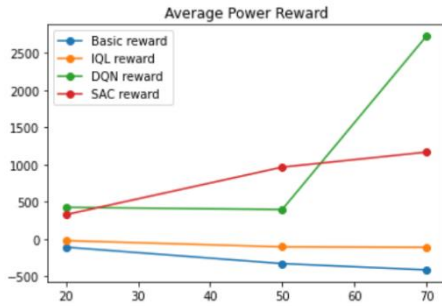


Εικόνα 18: Συνολική επιβράβευση σε πείραμα μεγαλύτερης κατανομής

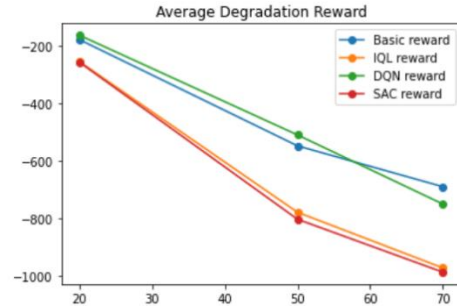
Για άλλη μια φορά το καλύτερο αποτέλεσμα είναι αυτό που υπολογίζει ο αλγόριθμος DQN, με τον αλγόριθμο SAC να μην απέχει πολύ. Πλέον ο αλγόριθμος IQL εμφανίζει και αυτός αποτέλεσμα καλύτερο από τον βασικό αλγόριθμο. Όλοι οι αλγόριθμοι προκαλούν λιγότερη φθορά στην μπαταρία του αυτοκινήτου από τον βασικό. Μπορούμε να συμπεράνουμε επομένως ότι στην περίπτωση που το πάρκινγκ μένει γεμάτο για μεγαλύτερο χρονικό διάστημα τα αποτελέσματα όλων των αλγορίθμων βελτιώνονται.

Πείραμα κατανομής Λιγότερων Αυτοκινήτων

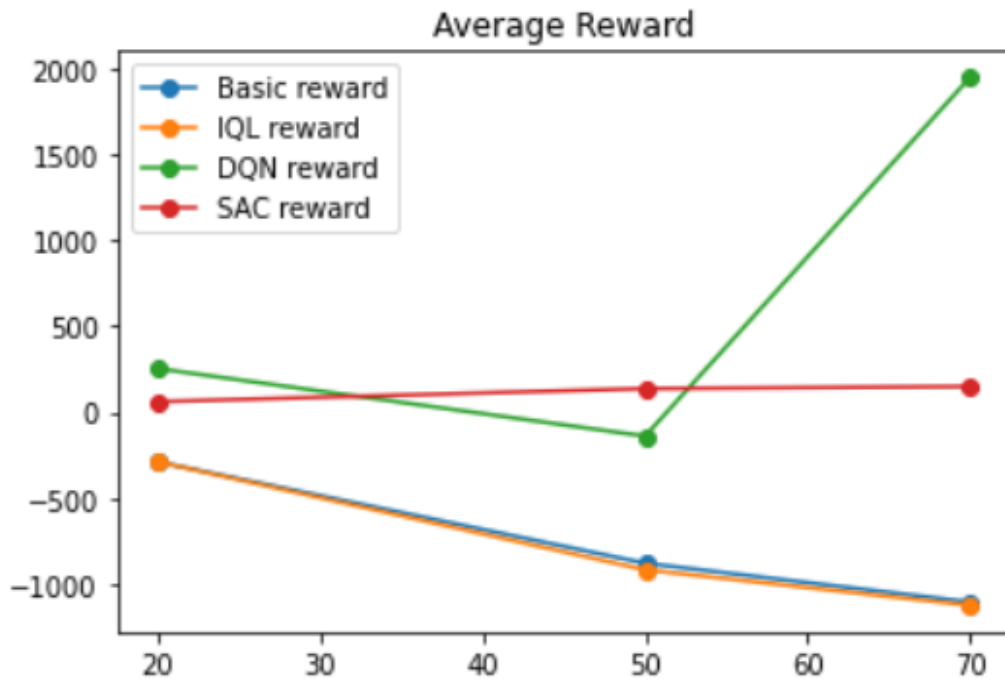
Τέλος οι αλγόριθμοι δοκιμάστηκαν σε κατανομή άφιξης αυτοκινήτων με ρυθμό άφιξης περίπου τον μισό από αυτόν της κατανομής εκπαίδευσης των νευρωνικών.



Εικόνα 19: Επιβράβευση λόγω αγοράς ηλεκτρικού ρεύματος σε πείραμα μικρότερης κατανομής



Εικόνα 20: Επιβράβευση λόγω φθοράς της μπαταρίας σε πείραμα μικρότερης κατανομής



Εικόνα 21: Συνολική επιβράβευση σε πείραμα μικρότερης κατανομής

Παρατηρούμε αποτελέσματα αρκετά παρόμοια με αυτά του πρώτου πειράματος. Οι αλγόριθμοι με το καλύτερο συνολικό αποτέλεσμα είναι οι DQN και SAC ενώ ο αλγόριθμος IQL για άλλη μια φορά δεν κατάφερε να βγάλει καλύτερο αποτέλεσμα από τον βασικό αλγόριθμο. Βλέπουμε όμως ότι στην περίπτωση του πάρκινγκ 50 θέσεων ο SAC ξεπέρασε τον DQN με μικρή διαφορά. Παρόλα αυτά ο DQN έχει πολύ καλύτερα αποτελέσματα στη περίπτωση του πάρκινγκ 70 θέσεων επομένως μπορούμε να συμπεράνουμε ότι έχει

καλύτερη κλιμακωσιμότητα. Όσο αφορά την φθορά της μπαταρίας ο DQN πέτυχε πολύ καλό αποτέλεσμα, όχι χειρότερο από αυτό του βασικού αλγορίθμου, ενώ οι αλγόριθμοι IQL και SAC προκάλεσαν περισσότερη φθορά με τις αποφάσεις τους.

Τελικά αποτελέσματα

Συνολικά μπορούμε να συμπεράνουμε ότι ο καλύτερος αλγόριθμος -δηλαδή αυτός που εμφανίζει τα καλύτερα αποτελέσματα στις περισσότερες περιπτώσεις- είναι ο αλγόριθμος DQN. Αυτό μπορεί να εξηγηθεί από την αποδοτικότητα των δειγμάτων εκπαίδευσης που εμφανίζει καθώς από όλους τους αλγορίθμους που δοκιμάστηκαν είναι αυτός που μπορεί να χρησιμοποιήσει καλύτερα τα δεδομένα που έχει αποκτήσει. Ο SAC εμφανίζει επίσης πολύ καλά αποτελέσματα αλλά όχι τόσο καλά όσο ο DQN. Μπορούμε να συμπεράνουμε λοιπόν ότι το πλεονέκτημα του SAC, δηλαδή το ότι το σύνολο αποφάσεων είναι συνεχές, μπορεί να αντισταθμιστεί από το μεγάλο πλήθος δυνατών αποφάσεων του DQN.

Ο αλγόριθμος IQL δεν κατάφερε να βρει κάποια καλή λύση στις περισσότερες περιπτώσεις πράγμα που μπορεί να εξηγηθεί από τα χαρακτηριστικά του προβλήματος. Συγκεκριμένα το πρόβλημα μπορεί να φαίνεται πως εμφανίζει χαρακτηριστικά προβλήματος πολλών πρακτόρων αλλά στην πραγματικότητα όλοι οι πράκτορες είναι ισοδύναμοι. Η διαφορά ανάμεσα σε 2 θέσεις πάρκινγκ είναι πρακτικά ανύπαρκτη αφού το κάθε αυτοκίνητο μπορεί να μπει σε οποιαδήποτε θέση. Επομένως όλοι οι πράκτορες για κάθε θέση είναι λογικό να έχουν τα ίδια βάρη αφού οι αποφάσεις

τους είναι ισοδύναμες. Συνεπώς η ύπαρξη πολλών πρακτόρων που προσπαθούν να εκπαιδεύσουν τα βάρη τους προσθέτει περιττή πολυπλοκότητα στην μάθηση ενώ ένας πράκτορας που παίρνει αποφάσεις και εκπαιδεύεται σε όλες τις θέσεις εμφανίζει καλύτερα αποτελέσματα.

Συνολικά μπορούμε να συμπεράνουμε ότι η χρήση αλγορίθμων μηχανικής μάθησης είναι ένας πολύ καλός τρόπος να λυθεί το πρόβλημα απόφασης του συσσωρευτή αφού τα αποτελέσματα στα οποία μπορεί να καταλήξει είναι σημαντικά καλύτερα από αυτά ενός βασικού αλγορίθμου. Μπορούμε επίσης να συμπεράνουμε ότι η πρακτική V2G με χρήση μηχανικής μάθησης μπορεί να επιφέρει σημαντικό κέρδος στον συσσωρευτή χωρίς να προκαλέσει ιδιαίτερη φθορά στα ηλεκτρικά αυτοκίνητα, συγκριτικά με τη απλή φόρτιση. Τέλος μπορούμε να συμπεράνουμε ότι αυτή η συμπεριφορά παραμένει ακόμα και σε περιπτώσεις στις οποίες τα χαρακτηριστικά του προβλήματος, δηλαδή η κατανομή αφίξεων και αποχωρήσεων των αυτοκινήτων, διαφέρει από την αναμενόμενη.

Βιβλιογραφία

1. Climate Impacts on Ecosystems. Ανάκτηση από EPA:
https://19january2017snapshot.epa.gov/climate-impacts/climate-impacts-ecosystems_.html
2. Climate Effects on Health. (2021, March 2). Ανάκτηση από CDC:
<https://www.cdc.gov/climateandhealth/effects/default.htm>
3. How is climate change affecting the economy and society? Ανάκτηση από IBERDROLA:
<https://www.iberdrola.com/environment/impacts-of-climate-change>
4. Energy and climate change. (2021 May 11). Ανάκτηση από European Environment Agency: <https://www.eea.europa.eu/signals/signals-2017/articles/energy-and-climate-change>
5. A European green Deal. Ανάκτηση από Europa.eu:
https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_el
6. Energy policy: general principles. (2021 May). Ανάκτηση από europarl.europa.eu:
<https://www.europarl.europa.eu/factsheets/en/sheet/68/energy-policy-general-principles>
7. What is Renewable Energy. Ανάκτηση από CEF:
<https://www.conserve-energy-future.com/advantages-and-disadvantages-of-renewable-energy.php#2> Renewable Energy Can be Unreliable
8. Janina C. Ketterer (20120) The impact of Wind Power Generation on the Electricity Price in Germany. Ανάκτηση από:
<https://www.econstor.eu/bitstream/10419/73830/1/lfoWorkingPaper-143.pdf>
9. Vehicle-to-grid. Ανάκτηση από Wikipedia: <https://en.wikipedia.org/wiki/Vehicle-to-grid>
10. Science for Environmental Policy (2019 March 27) Ανάκτηση από Europa.eu:
https://ec.europa.eu/environment/integration/research/newsalert/pdf/understanding_degradation_battery_life_key_successful_v2g_523na1_en.pdf
11. Reinforcement Learning. Ανάκτηση από Wikipedia:
https://en.wikipedia.org/wiki/Reinforcement_learning
12. Climate Change. Ανάκτηση από national Geographic:
<https://www.nationalgeographic.org/encyclopedia/climate-change/>
13. The causes of Climate Change. Ανάκτηση από Nasa: <https://climate.nasa.gov/causes/>
14. Myles Allen, Mustafa Babiker et al. (2018). Summary for Policymakers. Ανάκτηση από:
https://www.ipcc.ch/site/assets/uploads/sites/2/2019/05/SR15_SPM_version_report_HR.pdf
15. Dr Katherine Kramer. (2019). Counting the cost 2019: a year of climate breakdown. Ανάκτηση από: <https://www.christianaid.org.uk/sites/default/files/2019-12/Counting-the-cost-2019-report-embargoed-27Dec19.pdf>
16. Jonathan M.Harris, Brian Roach et al. (2017). The Economics of Global Climate Change. Ανάκτηση από:
https://www.bu.edu/eci/files/2019/06/The_Economics_of_Global_Climate_Change.pdf
17. The Global Risks Report 2021. (2021). Ανάκτηση από World Economic Forum:
http://reports.weforum.org/global-risks-report-2021/?doing_wp_cron=1627818792.1403870582580566406250

18. Diarmid Cambell-Lendrum, Dave D. Chadee et al. (2014). Human Health: Impacts, Adaptation and Co-Benefits. Ανάκτηση από:
https://www.ipcc.ch/site/assets/uploads/2018/02/WGIIAR5-Chap11_FINAL.pdf
19. Nick Watts, Markus Amann et al. (2019 November 13). The 2019 report of the Lancet Countdown on health and climate change: ensuring that the health of a child born today is not defined by a changing climate. Ανάκτηση από:
[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(19\)32596-6/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)32596-6/fulltext)
20. Quantitative risk assessment of the effects of climate change on selected causes of death, 2030s and 2050s. (2014). Ανάκτηση από World Health Organization:
https://apps.who.int/iris/bitstream/handle/10665/134014/9789241507691_eng.pdf
21. Dr Marco Springmann, Daniel Masson-D’Croz et al. (2016 March 2). Global and regional health effects of future food production under climate change: a modelling study. Ανάκτηση από: <https://www.thelancet.com/action/showPdf?pii=S0140-6736%2815%2901156-3>
22. Pachauri R. K. Meyer L.A. et al. (2014). Contribution of Working Groups I, II, and III to the Fifth Assessment Report of the intergovernmental Panel on Climate Change. Ανάκτηση από: <https://www.who.int/globalchange/global-campaign/cop21/en/>
23. Usonobong F. Akpan, Godwin E.Akpan. (2012). The Contribution of Energy Consumption to Climate Change: A Feasible Policy Direction. Ανάκτηση από:
<https://www.econjournals.com/index.php/ijeep/article/view/96/67>
24. Renewables can reduce CO2 emissions by 70% by 2050. (2017 march 21). Ανάκτηση από climate Action:
[https://www.climateaction.org/news/renewables can reduce co2 emission by 70 by 2050](https://www.climateaction.org/news/renewables%20can%20reduce%20co2%20emission%20by%2070%20by%202050)
25. Delivering the European Green Deal. Ανάκτηση από Europa.eu
https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal/delivering-european-green-deal_en
26. European Green Deal. Ανάκτηση από Wikipedia
https://en.wikipedia.org/wiki/European_Green_Deal
27. Renewable energy in the EU in 2018. (2020 January 23). Ανάκτηση από Europa.eu:
<https://ec.europa.eu/eurostat/documents/2995521/10335438/8-23012020-AP-EN.pdf/292cf2e5-8870-4525-7ad7-188864ba0c29>
28. Omar Ellabban, Haithan Abu-Rub, Frede Blaabjerg. (2014 November). Renewable energy Sources: Current status, future prospects and their enabling technology. Ανάκτηση από:
<https://www.sciencedirect.com/science/article/abs/pii/S1364032114005656?via%3Dihub>
29. Kerry Thoubboron. (2018 October 25). Advantages and disadvantages of renewable energy. Ανάκτηση από: <https://news.energysage.com/advantages-and-disadvantages-of-renewable-energy/>
30. Michael Shellenberger. (2019 April 22). Unreliable Nature of Solar And Wind Makes Electricity More Expensive, New Study Finds. Ανάκτηση από:
<https://www.forbes.com/sites/michaelshellenberger/2019/04/22/unreliable-nature-of-solar-and-wind-makes-electricity-much-more-expensive-major-new-study-finds/>

31. Benefits of Renewable Energy Use. (2017 December 20). Ανάκτηση από Union of Concerned Scientists: <https://www.ucsusa.org/resources/benefits-renewable-energy-use>
32. Renewable energy can power the world. (2011 May 9). Ανάκτηση από The Guardian: <https://www.theguardian.com/environment/2011/may/09/ipcc-renewable-energy-power-world>
33. Renewable Energy Now Accounts for a Third of Global Power Capacity. (2019 April 2). Ανάκτηση από IRENA: <https://www.irena.org/newsroom/pressreleases/2019/Apr/Renewable-Energy-Now-Accounts-for-a-Third-of-Global-Power-Capacity>
34. RENEWABLES GLOBAL FUTURES REPORT. (2017). Ανάκτηση από: https://www.ren21.net/wp-content/uploads/2019/06/GFR-Full-Report-2017_webversion_3.pdf
35. B.V.Mathiesen, H.Lund et al. (2015 May 1). Smart Energy Systems for coherent 100% renewable energy and transport solutions. Ανάκτηση από: <https://www.sciencedirect.com/science/article/abs/pii/S0306261915001117?via%3Dihub>
36. Understanding electricity markets in the EU. (2016 november). Ανάκτηση από: <http://www.easyres-project.eu/wp-content/uploads/2019/02/understanding-elecricity-markets-in-the-eu.pdf>
37. Knaut Andreas, Paulus Simon. (2016). Hourly price elasticity pattern of electricity demand in the German day-ahead market. Ανάκτηση από: <https://www.econstor.eu/bitstream/10419/144865/1/865176043.pdf>
38. Liutauras Vaitkus. (2014 January). Electricity market price during 24h in winter. Ανάκτηση από: https://www.researchgate.net/figure/Electricity-price-diagram-during-typical-winter-day-in-Lithuania_fig1_269224396
39. Hourly electricity consumption varies throughout the day and across seasons. (2020 February 21). Ανάκτηση από: <https://www.eia.gov/todayinenergy/detail.php?id=42915>
40. Eric P. Johnson, Matthew E.Oliver. (2016). Renewable Energy and wholesale Electricity Price Variability . ανάκτηση από: <https://www.iaee.org/en/publications/newsletterdl.aspx?id=323>
41. A123 Inks Deal to Develop Battery Cells for Gm Electric Car. (2007). Ανάκτηση από: <http://www.xconomy.com/boston/2007/08/10/a123-inks-deal-to-develop-battery-cells-for-gm-electric-car/>
42. Jim Gorzelany. Electric-Vehicle Battery Basics. Ανάκτηση από: <https://www.myeve.com/research/ev-101/electric-vehicle-battery-basics>
43. Barre, Harold (1997). Managing 12 Volts: How To Upgrade, Operate, and Troubleshoot 12 Volt Electrical Systemms.
44. Sandra Wappelhorst, Dale Hall, et al. (2020). Analyzing Policies to Grow the Electric Vehicle Market In European Cities. Ανάκτηση από: https://theicct.org/sites/default/files/publications/EV_city_policies_white_paper_fv_20_200224.pdf
45. Is my car bad for the environment? A guide to vehicle exhaust emissions. Ανάκτηση από: <https://www.rac.co.uk/drive/advice/emissions/vehicle-exhaust-emissions-what-comes-out-of-your-cars-exhaust/>

46. Energy Density. Ανάκτηση από Wikipedia:
https://en.wikipedia.org/w/index.php?title=Energy_density&oldid=949095221
47. Toyota RAV4 EV. Ανάκτηση από Wikipedia
https://en.wikipedia.org/w/index.php?title=Toyota_RAV4_EV&oldid=952112719
48. Charlotte Argue. (2020 July 7). What can 6000 Electric Vehicles tell us about EV battery health?. Ανάκτηση από: <https://www.geotab.com/blog/ev-battery-health/>
49. Sivapriya M.Bhagavathy, Hannah Budnitz et al. (2021 March 23). Impact of Charging Rates on Electric Vehicle Battery Life. Ανάκτηση από:
<https://findingspress.org/article/21459-impact-of-charging-rates-on-electric-vehicle-battery-life>
50. Anderson Hoke, Alexander Brissette, et al. (2011). Electric vehicle charge optimization including effects of lithium-ion battery degradation. Ανάκτηση από:
<https://ieeexplore.ieee.org/abstract/document/6043046>
51. Jacob Sorme. (2020). Intelligent Charging Algorithm for Electric Vehicles. Ανάκτηση από:
<http://www.diva-portal.org/smash/get/diva2:1466882/FULLTEXT01.pdf>
52. Anderson Hoke, Alexander Brissette, et al. (2011). Electric vehicle charge optimization including effects of lithium-ion battery degradation. Ανάκτηση από:
<https://ieeexplore.ieee.org/abstract/document/6043046>
53. Yongquan sun, Saurabh Saxena, Michael Pecht, (2018). Derating Guidelines for Lithium-Ion Batteries. Ανάκτηση από:
https://www.researchgate.net/publication/329212716_Derating_Guidelines_for_Lithium-Ion_Batteries
54. Sivapriya M. Bhagavathy, Hannah Budnitz et al. (2021). Impact of Charging rates on Electric Vehicle Battery Life. Ανάκτηση από: <https://findingspress.org/article/21459.pdf>
55. Willet Kempton, Steven E. Letendre. (1996). Electric Vehicles As a new Power Source for Electric Utilities. Ανάκτηση από:
<http://www.civil.ist.utl.pt/~martinez/PDF/ELECTRICMOVE/Paper18.pdf>
56. Habibalah Kh. Khanekhdani, Masoud M. Tafreshi, and Marzieh Khosravi. (2013). Modeling operation of electric vehicles aggregator in reserve services market by using game theory method. Ανάκτηση από:
<https://aip.scitation.org/doi/10.1063/1.4850524>
57. Jasna Tomic. Willet Kempton. (2007). Using fleets of electric-drive vehicles for grid support. Ανάκτηση από:
https://www.researchgate.net/publication/222651706_Using_fleets_of_electric-drive_vehicles_for_grid_support
58. Diyun Wu. K.T. Chau. (2012). Transient Stability Analysis of SMES for Smart Grid with Vehicle-to-Grid
https://www.researchgate.net/publication/254060355_Transient_Stability_Analysis_of_SMES_for_Smart_Grid_With_Vehicle-to-Grid_Operation
59. Murat Yilmaz, Philip T. Krein et al. (2012). Review of benefits and challenges of vehicle-to-grid technology. Ανάκτηση από: <https://ieeexplore.ieee.org/document/6342356>
60. Brooks A.N. (2002). Vehicle-to-grid Demonstration Project grid regulation ancillary service with a battery electric vehicle. Ανάκτηση από: <https://trid.trb.org/view/661306>
61. Load management. Ανάκτηση από Wikipedia:
https://en.wikipedia.org/wiki/Load_management

-
62. Kristien Clement-Nyns, Edwin Haesen and Johan Driesen. (2011). The impact of vehicle-to-grid on the distribution grid. Ανάκτηση από:
<http://www.civil.ist.utl.pt/~martinez/PDF/ELECTRICMOVE/Paper22.pdf>
 63. Christophe Guille and George Gross. (2009). A conceptual framework for the vehicle-to-grid (V2G) implementation
https://econpapers.repec.org/article/eeeeenepol/v_3a37_3ay_3a2009_3ai_3a11_3ap_3a4379-4390.htm
 64. M. J. Ruthford and V. Yousefzadeh. (2011). The Impact of Electric Vehicle Battery Charging on Distribution Transformers.
 65. S. S. Raghavan and A. Khaligh, (2012 January). Impact of Plug-in Hybrid Electric Vehicle Charging on a Distribution Network in a Smart Grid Environment
 66. J. Lassila, J. Haakana, V. Tikka, and J. Partanen. (2012 March). Methodology to Analyze the Economic Effects of Electric Cars as Energy Storages
 67. Francisco S. Melo. Convergence of Q-Learning: A Simple Proof. Ανάκτηση από:
<http://users.isr.ist.utl.pt/~mtjspaan/readingGroup/ProofQlearning.pdf>
 68. Reinforcement Learning. Ανάκτηση από:
<https://www.cse.unsw.edu.au/~cs9417ml/RL1/algorithms.html>
 69. Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction. Ανάκτηση από: <http://incompleteideas.net/sutton/book/ebook/the-book.html>
 70. Vincent Francois-Lavet. (2015). How to Discount Deep Reinforcement Learning: Towards New Dynamic Strategies. Ανάκτηση από: <https://arxiv.org/abs/1512.02011>
 71. Optimistic Initial Values. Ανάκτηση από:
<https://web.archive.org/web/20130908031737/http://webdocs.cs.ualberta.ca/~sutton/book/ebook/node21.html>
 72. Kung-Hsiang, Huang. (2018 January 12). Introduction to Various Reinforcement Learning Algorithms. Ανάκτηση από: <https://towardsdatascience.com/introduction-to-various-reinforcement-learning-algorithms-i-q-learning-sarsa-dqn-ddpg-72a5e0cb6287>
 73. Jacob Foerster, Nantas Nardelli et al. (2018). Stabilising Experience Replay for Deep Multi-Agent Reinforcement Learning. <https://arxiv.org/pdf/1702.08887.pdf>
 74. Kamil Andrzej Ciosek, Shimon Whiteson. (2017). Off-Environment Reinforcement Learning. Ανάκτηση από:
<https://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14378>
 75. Jan Peters, J. Andrew Bagnell. Policy Gradient Methods. Ανάκτηση από:https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-30164-8_640
 76. Chris Yoon. (2018). Deriving Policy Gradients and Implementing Reinforce. Ανάκτηση από: <https://medium.com/@thechrisyoon/deriving-policy-gradients-and-implementing-reinforce-f887949bd63>
 77. Chris Yoon. (2018). Understanding Actor Critic Methods and A2C. Ανάκτηση από:
<https://towardsdatascience.com/understanding-actor-critic-methods-931b97b6df3f>
 78. Tuomas Haarnoja, Aurick Zhou et al. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a stochastic Actor. Ανάκτηση από:
<https://proceedings.mlr.press/v80/haarnoja18b/haarnoja18b.pdf>
 79. SOFT Actor-Critic. Ανάκτηση από:
<https://spinningup.openai.com/en/latest/algorithms/sac.html>
 80. Yonquan Sun, Saurabh Saxena, Michael Pecht. (2018). Derating Guidelines for Lithium-Ion Batteries. Ανάκτηση από: <https://www.mdpi.com/1996-1073/11/12/3295/pdf/1>
-

81. Sources of Greenhouse Gas Emissions. Ανάκτηση από EPA:
<https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions>