ΣΤΡΑΤΙΩΤΙΚΗ ΣΧΟΛΗ ΕΥΕΛΠΙΔΩΝ
Τμήμα Στρατιωτικών Επιστημών

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΔΙΙΔΡΥΜΑΤΙΚΟ ΔΙΑΤΜΗΜΑΤΙΚΟ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ
ΣΠΟΥΔΩΝ ΑΚΑΔΗΜΑΪΚΟΥ ΕΤΟΥΣ 2024
ΣΧΕΔΙΑΣΗ&ΕΠΕΞΕΡΓΑΣΙΑ
ΣΥΣΤΗΜΑΤΩΝ (SYSTEMS ENGINEERING)

(ΠΔ 96/2015/ΦΕΚ163Α'/20.08.2014)

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
Σχολή Μηχανικών Παραγωγής & Διοίκησης

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΜΕΛΕΤΗ ΠΡΟΒΛΗΜΑΤΩΝ

ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ

ΜΕ ΤΗ ΧΡΗΣΗ ΜΑΡΚΟΒΙΑΝΩΝ

ΔΙΑΔΙΚΑΣΙΩΝ ΑΠΟΦΑΣΕΩΝ

Διατριβή που υπεβλήθη για την μερική ικανοποίηση των απαιτήσεων για την απόκτηση Μεταπτυχιακού Διπλώματος Ειδίκευσης

Υπό:

ΑΛΕΞΑΝΔΡΟΥ ΚΟΝΤΟΓΕΩΡΓΑΚΗ
Α.Μ.:2018018005

ΟΚΤΩΒΡΙΟΣ 2024

Η Μεταπτυχιακή Διατριβή του ΑΛΕΞΑΝΔΡΟΥ ΚΟΝΤΟΓΕΩΡΓΑΚΗ εγκρίνεται:

ΤΡΙΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Αναπληρωτής Καθηγητής **Καραματσούκης Κωνσταντίνος (Επιβλέπων)** Κ. Καραματσούκης

Καθηγητής **Νικόλαος Δάρας**

Αναπληρωτής Καθηγητής **Στέλιος Τσαφαράκης** Stelios Tsafarakis Digitally signed by Stelios Tsafarakis DN: c=GR, l=Chania, o=Technical University of Crete, sn=Tsafarakis, givenName=Stelios, serialNumber=4877168308, cn=Stelios Tsafarakis Date: 2024.10.31 10:00:38 +02'00'

Έτος 2024

ΕΥΧΑΡΙΣΤΙΕΣ

ΕΝ ΟΝΟΜΑΤΙ ΤΗΣ ΑΓΙΑΣ ΚΑΙ ΟΜΟΟΥΣΙΑΣ ΚΑΙ ΑΔΙΑΙΡΕΤΟΥ
ΤΡΙΑΔΟΣ

Ευχαριστώ τον Όσιο Ιάκωβο τον Εν Ευβοίας όπου άντλησα το κουράγιο προκειμένου να πραγματοποιηθεί αυτή η εργασία και φυσικά θα ήθελα να ευχαριστήσω τον καθηγητή μου κ. Κωνσταντίνο Καραματσούκη για την υπομονή και το χρόνο που διέθεσε για να την ολοκληρώσω.

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Έχω διαβάσει και κατανοήσει τους κανόνες για τη λογοκλοπή και τον τρόπο σωστής αναφοράς των πηγών που περιέχονται στον Κανονισμό Σπουδών του Δ.Π.Μ.Σ. της ΣΣΕ και της Σχολής Μηχανικών Παραγωγής και Διοίκησης του Πολυτεχνείου Κρήτης. Δηλώνω ότι, από όσα γνωρίζω, το περιεχόμενο της παρούσας Διπλωματικής Εργασίας είναι προϊόν δικής μου δουλειάς και υπάρχουν αναφορές σε όλες τις πηγές που χρησιμοποίησα.

Ημερομηνία

Ο Δηλών

11 Οκτωβρίου 2024

Αλέξανδρος Κοντογεωργάκης

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΙΝΑΚΑΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

ΠΙΝΑΚΑΣ ΠΙΝΑΚΩΝ

ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ

ΠΕΡΙΛΗΨΗ

Τα τελευταία χρόνια υπήρξε μια τεράστια ανάπτυξη στον τομέα της τεχνολογίας. Πολλές εφαρμογές έχουν αντικατασταθεί από την τεχνολογία, η οποία έχει βοηθήσει σημαντικά στην εξέλιξη και στην βελτίωση των συνθηκών του ανθρώπου. Η Τεχνητή Νοημοσύνη (Artificial Intelligence) είναι μια από τις τεχνολογικές καινοτομίες που συνέβησαν, για να αντικαταστήσουν την επέμβαση του ανθρώπου σε πολλές δραστηριότητες και διαδικασίες λήψης αποφάσεων. Με τον όρο τεχνητή νοημοσύνη, θεωρείται ο τομέα της επιστήμης των υπολογιστών που ασχολείται με τον σχεδιασμό συστημάτων καθώς και την υλοποίηση που μπορούν να εξομοιώσουν γνωστικές ικανότητες του ανθρώπου σε πολύ μεγάλο βαθμό.

Στόχος της παρούσας μεταπτυχιακής διατριβής είναι η μοντελοποίηση και επίλυση προβλημάτων της Τεχνητής Νοημοσύνης στο πλαίσιο που παρέχουν οι Μαρκοβιανές Διαδικασίες Αποφάσεων. Αρχικά, θα παρουσιαστούν κάποιες βασικές έννοιες της Τεχνητής Νοημοσύνης (Artificial Intelligence) καθώς και της Ενισχυτικής μάθησης (Reinforcement Learning). Στην συνέχεια θα παρουσιαστούν οι βασικές έννοιες και οι αλγόριθμοι των Μαρκοβιανών Διαδικασιών Αποφάσεων. Θα μελετηθούν κάποιες εφαρμογές όπως για παράδειγμα η διαδικασία έρευνας και διάσωσης με ένα μη επανδρωμένο ελικόπτερο σε ένα άγνωστο περιβάλλον με σκοπό την εύρεση της βέλτιστης ζώνης προσγείωσης για την εκτέλεση της αποστολής του και θα παρουσιαστούν οι αλγόριθμοι που έχουν αναπτυχθεί για την επίλυση του παραπάνω προβλήματος.

Keywords: Markov Chain, Markov decision Process, Reinforcement Learning, UAV, Artificial Intelligence, Military Application

ΚΕΦΑΛΑΙΟ 1

Τεχνητή Νοημοσύνη

1.1 Τεχνητή Νοημοσύνη

Όπως ανέφεραν οι Poole, Macworth και Goebel (1998), η τεχνητή νοημοσύνη (AI) είναι ένας συνδυασμός επιστήμης των υπολογιστών και της φυσιολογίας . Πιο απλά, η νοημοσύνη είναι το υπολογιστικό μέρος της ικανότητας επίτευξης στόχων παγκοσμίως. Η νοημοσύνη είναι η ικανότητα σκέψης, φαντασίας, δημιουργίας γνώσης και κατανόησης, αναγνώρισης προτύπων, προσαρμογής στην αλλαγή και επιλογής μάθησης από εμπειρίες. Ο άνθρωπος έχει την τάση να μοιάζει και να μιμείται το είδος του δηλαδή τον άνθρωπο. Άρα όπως αναφέρθηκε προηγουμένως η τεχνητή νοημοσύνη κάνει τους υπολογιστές να μοιάζουν άνθρωποι περισσότερο από τον ίδιο τον άνθρωπο και μάλιστα πολύ πιο γρήγορα. (Chhaya & Sarode, 2020)

1.2 Μάθηση

Η μάθηση ορίζεται ως η απόκτηση γνώσεων ή δεξιοτήτων, σε έναν συγκεκριμένο τομέα. Ο ορισμός αυτός αφορά τον άνθρωπο. Στην ψυχολογία, γενικευμένοι ορισμοί της μάθησης και πολλοί από αυτούς ερμηνεύουν τη μάθηση ως τροποποίηση συμπεριφοράς δεδομένης μιας κατάστασης, ή ως ακολουθία των επαναλαμβανόμενων εμπειριών. (Chhaya & Sarode, 2020)

1.3 Μηχανική Μάθηση

Αναφέρεται (Bavakutty & Haneefa (2006)), ότι η μάθηση σημαίνει απόκτηση νέων γνώσεων για τη βελτίωση των δεξιοτήτων ενός ατόμου. Η βελτίωση διαφόρων πτυχών της διαδικασίας, όπως η απόκτηση γνώσης, η απόκτηση βασικής κατανόησης, η κατανόηση του θέματος και των αντίστοιχων αντιδράσεων και η κατανόηση της σχέσης μεταξύ των αντίστοιχων πεδίων, μπορεί να ερμηνευθεί βιολογικά ως ενίσχυση του πρωτοτύπου της νευρικής σύνδεσης για την εκτέλεση της επιθυμητής λειτουργίας.

Ο Bishop (2006) στην επιστημονική μελέτη αλγορίθμων και στατιστικών μοντέλων ονομάζει μηχανική μάθηση όταν χρησιμοποιούνται συστήματα υπολογιστών για την εκτέλεση ορισμένων εργασιών. Θεωρείται τμήμα της τεχνητής νοημοσύνης. Οι αλγόριθμοι μηχανικής μάθησης δημιουργούν μαθηματικά μοντέλα βασισμένα σε δείγματα δεδομένων, που ονομάζονται δεδομένα εκπαίδευσης, για να κάνουν προβλέψεις και αποφάσεις χωρίς να προγραμματίζονται ρητά για την πραγματοποίηση των στόχων τους. (Chhaya & Sarode, 2020)

1.4 Πεδία της Τεχνητής Νοημοσύνης

Η τεχνητή νοημοσύνη επικεντρώνεται σε συμβολικές, μη αλγοριθμικές λύσεις προβλημάτων. Η νοημοσύνη βασίζεται στην ικανότητα χειρισμού συμβόλων. Η Τεχνητή νοημοσύνη έχει αλλάξει την κοινωνία πέρα από τη φαντασία. Οι στόχοι των υποτομέων του έμπειρα συστήματα, Επεξεργασία Φυσικής Γλώσσας, αναγνώριση προτύπων και η ρομποτική είναι μια προσομοίωση της ανθρώπινης νοημοσύνης χρησιμοποιώντας έναν υπολογιστή. Ορισμένα πεδία που χρησιμοποιούνται στις τελευταίες τεχνολογίες και την ανάπτυξη υπολογιστών παρατίθεται στον πίνακα (Chhaya & Sarode, (2020))

Πίνακας 1. Πεδία της Τεχνητή Νοημοσύνη

Τεχνητή Νοημοσύνη	Εφαρμογές Γνωστικής επιστήμης	Εξειδικευμένα Συστήματα Συστήματα Μάθησης Ασαφής λογική Γενετικοί Αλγόριθμοι Ουδέτερα Δίκτυα Ευφυείς Πράκτορες
	Εφαρμογές Ρομποτικής	Οπτικές Αντιλήψεις Απτική Επιδεξιότητα Μετακίνηση Πλοήγηση
		Φυσικές Γλώσσες

	Εφαρμογές Φυσικής Διεπαφής	Αναγνωρίσεις ομιλίας Πολυαισθητηριακές διεπαφές Εικονική πραγματικότητα

(Chhaya & Sarode, 2020)

1.5 Εφαρμογές της Τεχνητής Νοημοσύνη

Οι υπολογιστές παρέχουν το τέλειο περιβάλλον για να πειραματιστούν και να εφαρμόσουν την τεχνητή νοημοσύνη. Η τεχνητή νοημοσύνη έχει μεγαλύτερη επιτυχία σε πνευματικές εργασίες όπως παιχνίδια που βασίζονται στους υπολογιστές και αποδείξεις θεωρημάτων παρά σε εργασίες που απαιτούν αντίληψη. Αυτά τα προγράμματα υπολογιστών έχουν στόχο την ενθάρρυνση της ανθρώπινης συμπεριφοράς και έχουν επίσης γίνει τεχνικές εφαρμογές όπως δημιουργία με τη βοήθεια υπολογιστή (CAI).

Συνήθως ο κύριος στόχος είναι να βρεθεί οποιαδήποτε τεχνική που εκτελεί γρήγορα τις διεργασίες με τον καλύτερο τρόπο. Η εφαρμογή της τεχνητής νοημοσύνης σχετίζεται με την εφαρμογή ευφυών συστημάτων και φυσικά της ρομποτικής. (Chhaya & Sarode, (2020))

1.6 Η Ενισχυτική Μάθηση

Η **ενισχυτική μάθηση** *(reinforcement learning)* στην επιστήμη των υπολογιστών είναι ένας γενικός όρος όπου το σύστημα μάθησης επιχειρεί να μάθει μέσα από την άμεση αλληλεπίδραση με το περιβάλλον. Εφαρμόζεται στον έλεγχο κινήσεων ρομπότ, στη βελτιστοποίηση εργασιών σε εργοστάσια, στη μάθηση επιτραπέζιων παιχνιδιών, κτλ. Η ενισχυτική μάθηση εμπνέεται από μαθησιακές έννοιες όπως είναι τα επιτραπέζια παιχνίδια με επιβράβευση και τιμωρία που εμφανίζονται ως πρότυπο. Σκοπός του συστήματος μάθησης είναι να μεγιστοποιήσει τη συνάρτηση ανταμοιβής. Το σύστημα πρέπει να ανακαλύψει μόνο του ποιες ενέργειες είναι αυτές που θα του αποφέρουν το μεγαλύτερο κέρδος.

Η ιδέα ότι η μάθηση πραγματοποιείται μέσω της αλληλεπίδρασης με το περιβάλλον είναι θεμελιώδης για την κατανόηση της μαθησιακής διαδικασίας. Μέσω της συνεχούς αλληλεπίδρασης

με το περιβάλλον, πραγματοποιείται συσσώρευση γνώσης για την επίτευξη στόχων και πλοήγησης σ' αυτό. Η κύρια πηγή μάθησης είναι οι αλληλεπιδράσεις με το περιβάλλον. Είτε αποκτούμε πρακτικές δεξιότητες όπως η ποδηλασία (ισορροπία χρήση του ποδηλάτου) είτε κοινωνικές δεξιότητες όπως η διαχείριση του χρόνου, βασιζόμαστε στην ανατροφοδότηση από το περιβάλλον μας για να καθοδηγήσουμε τη συμπεριφορά μας. Η διαδικασία μάθησης από την αλληλεπίδραση με το περιβάλλον συνδέεται άρρηκτα με θεωρίες μάθησης και νοημοσύνης τονίζοντας τη σημασία της στην ανθρώπινη ανάπτυξη και προσαρμογή. Στον πυρήνα της, η ιδέα της μάθησης μέσω της αλληλεπίδρασης διαμορφώνει την κατανόησή μας για το πώς αποκτούμε γνώση και προσαρμόζουμε σε διαφορετικές καταστάσεις καθ' όλη τη διάρκεια της ζωής μας, τονίζοντας τη δυναμική σχέση μεταξύ του ατόμου και του περιβάλλοντός του.

Η πλήρης προδιαγραφή των προβλημάτων ενίσχυσης μάθησης εστιάζει στον βέλτιστο έλεγχο των Μαρκοβιανών Διαδικασιών λήψης αποφάσεων, με βάση τις πιο σημαντικές πτυχές του πραγματικού προβλήματος. Κεντρική ιδέα είναι η αναπαράσταση του παράγοντα μάθησης που αλληλεπιδρά με το περιβάλλον του, με σκοπό την επίτευξη ενός καθορισμένου στόχου. (Sutton and Barto (2014, 2015))

ΚΕΦΑΛΑΙΟ 2

Εισαγωγή

2.1 Γενικές Πληροφορίες

Η παρούσα βιβλιογραφική εργασία πραγματεύεται προβλήματα απόφασης τα οποία ονομάζονται προβλήματα απόφασης υπό αβεβαιότητα. Αποτελούν κατηγορίες προβλημάτων της τεχνητής νοημοσύνης. Το βιβλίο των (Siguad & Buffet, 2010) πλαισιώνουν τις Μαρκοβιανές Διαδικασίες Αποφάσεων σε αποτελεσματικές τεχνικές επίλυσης όπως η Ενισχυτική μάθηση και οι Μαρκοβιανές Αλυσίδες.

Αρχικά παρουσιάζεται η θεωρία των Μαρκοβιανών Αλυσίδων Αποφάσεων καθώς και τη θεωρία που πλαισιώνει η Τεχνητή Νοημοσύνη. Στη συνέχεια παρουσιάζεται η εφαρμογή των Μαρκοβιανών Αλυσίδων Αποφάσεων σε ένα πρόβλημα βελτιστοποίησης στρατηγικής για ένα αυτόνομο ελικόπτερο έρευνας και διάσωσης που εξερευνά ζώνες προσγείωσης σε άγνωστο και αβέβαιο περιβάλλον. Η θεωρία των Μαρκοβιανών αλυσίδων είναι ένα βασικό μαθηματικό εργαλείο που χρησιμοποιείται για τη μοντελοποίηση της αβέβαιης και πιθανοτικής εξέλιξης συστημάτων.

Οι Μαρκοβιανές αλυσίδες χρησιμοποιούνται ευρέως σε πολλούς τομείς, συμπεριλαμβανομένων της πιθανοτικής μοντελοποίησης, της τηλεπικοινωνίας, της ρομποτικής, της επιστήμης των υπολογιστών και της οικονομίας (Siguad και Buffet, 2010)

Οι Μαρκοβιανές Αλυσίδες εμφανίσθηκαν στις αρχές της δεκαετίας του 1905. Το όνομα Markov αναφέρεται στον Ρώσο μαθηματικό Andrey Markov που έπαιξε καθοριστικό ρόλο στη διαμόρφωση των στοχαστικών διεργασιών.

Στις πρώτες μέρες τους, οι Μαρκοβιανές αλύσιδες ήταν γνωστό ότι έλυναν ζητήματα που σχετίζονται με τη διαχείριση και τον έλεγχο των αποθεμάτων, τη βελτιστοποίηση της ουράς και για θέματα καθορισμού διαδρομών.

Σήμερα, βρίσκουν εφαρμογές στη μελέτη προβλημάτων βελτιστοποίησης μέσω δυναμικού προγραμματισμού, ρομποτικής, αυτόματου ελέγχου, οικονομίας, κατασκευών κ.λπ

2.2 Η ανάλυση Markov

Η ανάλυση Markov αναφέρεται στον προσδιορισμό της πιθανότητας να συμβεί κάτι στο μέλλον, με βάση τις υπάρχουσες σήμερα πιθανότητες. Αν υποτεθεί ότι ένα σύστημα ξεκινάει από μια δεδομένη αρχική κατάσταση, ή συνθήκη, οι πιθανότητες να μεταβληθεί η κατάσταση του από την αρχική σε μια άλλη αναφέρονται στον πίνακα πιθανοτήτων μετάβασης. Η κατάσταση ενός συστήματος χρησιμοποιείται για να καθορίσει όλες τις πιθανές συνθήκες που εμφανίζονται σε μια διαδικασία ή ένα σύστημα. Η ανάλυση Markov υποθέτει ότι οι καταστάσεις που εμφανίζει ένα σύστημα είναι όλες όσες υπάρχουν και είναι επίσης αμοιβαία αποκλειόμενες. Σε πολλά προβλήματα η προσέγγιση μέσω Μαρκοβιανών στοχαστικών διαδικασιών θεωρεί τα παρακάτω:

- η πιθανότητα της μεταβολής μιας κατάστασης σε μια άλλη παραμένει διαχρονικά η ίδια
- να υπάρχει ένας περιορισμένος αριθμός πιθανών καταστάσεων
- είναι δυνατή η πρόγνωση κάθε μελλοντικής κατάστασης γνωρίζοντας την παρούσα κατάσταση και τον πίνακα πιθανοτήτων μετάβασης
- το μέγεθος και η εικόνα του συστήματος παραμένει αμετάβλητη κατά τη διάρκεια της ανάλυσης.

Αν πληρούνται όλες οι προϋποθέσεις, το επόμενο στάδιο της ανάλυσης είναι η εύρεση της πιθανότητας να βρίσκεται το σύστημα στη συγκεκριμένη κατάσταση.

Η πληροφορία για ποια κατάσταση βρίσκεται το σύστημα παρουσιάζεται μέσα από το ακόλουθο διάνυσμα πιθανοτήτων για κάθε κατάσταση, κατά την περίοδο i :

$\pi(i) = (\pi_1, \pi_2, \pi_3 \ldots\ldots, \pi_n)$ (η οποία καλείται στάσιμη κατανομή)

όπου n είναι ο αριθμός των καταστάσεων και $(\pi_1, \pi_2, \pi_3 \ldots\ldots, \pi_n)$ είναι η πιθανότητα να βρεθεί το σύστημα στη κατάσταση 1 , κατάσταση 2, …., κατάσταση n.

Στις περισσότερες περιπτώσεις όμως στο σύστημα υπάρχουν περισσότερα από ένα στοιχεία ή καταστάσεις και ως εκ τούτου είναι πολύπλοκο να υπολογισθεί το τι θα γίνει μετά την πρώτη

περίοδο χωρίς τη χρήση των πινάκων. Ο πίνακας πιθανοτήτων μετάβασης επιτρέπει ακριβώς αυτόν τον υπολογισμό των πιθανοτήτων μετάβασης από την παρούσα στη μελλοντική κατάσταση. Η πιθανότητα να βρεθεί το σύστημα, από την παρούσα κατάσταση i σε μια μελλοντική κατάσταση j, εμφανίζεται ως P_{ij} (για παράδειγμα P_{12} είναι η πιθανότητα να βρεθεί στη κατάσταση 2 ένα στοιχείο του οποίου η προηγούμενη κατάσταση ήταν η 1). Ο πίνακας P επομένως των πιθανοτήτων μετάβασης θα είναι ο ακόλουθος:

Πίνακας 2

Κατάσταση	**1**	**2**	**i**	**…**	**n**
1	$p_{1\to1}$	$p_{1\to2}$	$p_{1\to i}$	…	$p_{1\to n}$
2	$p_{2\to1}$	$p_{2\to2}$	$p_{2\to i}$	…	$p_{2\to n}$
i	$p_{i\to1}$	$p_{i\to1}$	$p_{i\to i}$	…	$p_{i\to n}$
…	…	…	…	…	…
n	$p_{n\to1}$	$p_{n\to1}$	$p_{n\to i}$	…	$p_{n\to n}$

Όπου όλες οι τιμές των P_{ij} καθορίζονται εμπειρικά και το άθροισμα με τις πιθανότητες σε μια στήλη θα πρέπει να ισούται με 1. Ξεκινώντας από την παρούσα περίοδο 0 οι πιθανότητες για την επόμενη περίοδο 1 θα δίνονται από τη σχέση $\pi(1) = \pi(0)*P$ και γενικότερα από κάθε περίοδο n είναι δυνατό να προσδιορίσουμε τις πιθανότητες των διαφόρων καταστάσεων για την περίοδο n+1 ως $\pi(n+1) = \pi(n)*P = \pi(0)*P_n$.

2.3 Πως λειτουργεί η Μαρκοβιανή Διαδικασία

Η Μαρκοβιανή Διαδικασία αναφέρεται σε μια στοχαστική διαδικασία λήψης αποφάσεων που χρησιμοποιεί ένα μαθηματικό πλαίσιο για να μοντελοποιήσει τη λήψη αποφάσεων για ένα δυναμικό σύστημα. Στα σενάρια που χρησιμοποιείται, τα αποτελέσματα είναι είτε τυχαία είτε λαμβάνονται διαδοχικές αποφάσεις με την πάροδο του χρόνου από τον υπεύθυνο λήψης αποφάσεων. Οι Μαρκοβιανές Αλυσίδες αξιολογούν ποιες ενέργειες πρέπει να λάβει ο υπεύθυνος λήψης αποφάσεων (λαμβάνοντας υπόψη την τρέχουσα κατάσταση και το περιβάλλον του συστήματος (Puterman, 1994). Σε μια Μαρκοβιανή Αλυσίδα (MDP) υπάρχει λοιπόν ένας λήπτης αποφάσεων, ένα σύνολο καταστάσεων S, ένα σύνολο ενεργειών A και ένα σύνολο ανταμοιβών R. Καθένα από τα σύνολα αυτά θεωρείται ότι περιέχει ένα πεπερασμένο αριθμό στοιχείων. Ο υπεύθυνος που λαμβάνει τις αποφάσεις αναφέρεται σε ένα σύστημα και είναι υπεύθυνος για τη λήψη αποφάσεων και την εκτέλεση ενεργειών. Λειτουργεί σε ένα περιβάλλον που περιγράφει λεπτομερώς τις διάφορες καταστάσεις στις οποίες αυτός βρίσκεται ενώ μεταβαίνει από τη μια κατάσταση στην άλλη. Η Μαρκοβιανή Διαδικασία Αποφάσεων ορίζει τον μηχανισμό με τον οποίο ορισμένες καταστάσεις και ενέργειες ενός λήπτης αποφάσεων οδηγούν στις επόμενες καταστάσεις. Επιπλέον, ο λήπτης αποφάσεων λαμβάνει ανταμοιβές ανάλογα με τη ενέργειας που εκτελεί και την κατάσταση που επιτυγχάνει (τρέχουσα κατάσταση). Η επιλογή για τα Μαρκοβιανά Μοντέλα Αποφάσεων αποκαλύπτει την ακόλουθη ενέργεια του λήπτη αποφάσεων ανάλογα με την τρέχουσα κατάστασή του.

Το μοντέλο χρησιμοποιεί την έννοια της Μαρκοβιανής ιδιότητας, η οποία δηλώνει ότι το μέλλον μπορεί να προσδιοριστεί μόνο μέσα από την παρούσα κατάσταση που περιλαμβάνει όλες τις απαραίτητες πληροφορίες από το παρελθόν (Norris, 1997). Μία Μαρκοβιανή διαδικασία εκφράζεται ως εξής:

$$P[S_{t+1}|S_t] = P[S_{t+1}|S_1, S_2, S_3 \ldots . S_t]$$

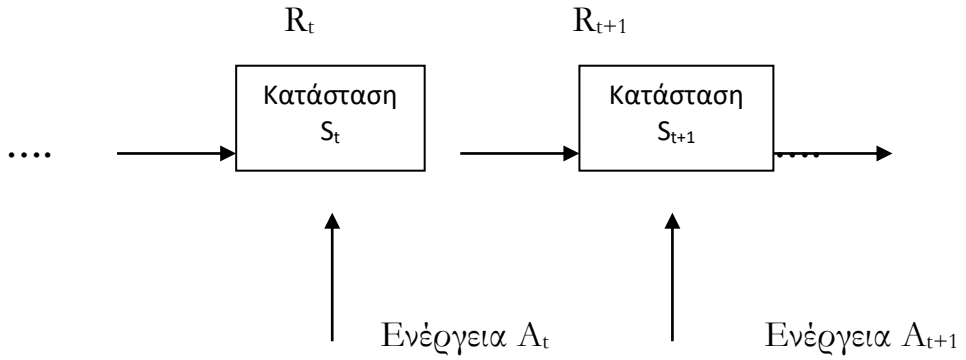
Σε κάθε χρονική στιγμή t= 0,1,2,3,…, ο λήπτης αποφάσεων δέχεται μια πληροφορία που αποτυπώνει μια συγκεκριμένη κατάσταση για το περιβάλλον, επιλεγμένη μέσα από το σύνολο S ($S_t \in S$). Βασισμένος στη κατάσταση αυτή ο πράκτορας επιλέγει μια ενέργεια A_t μέσα από το

σύνολο δράσεων Α. Το συγκεκριμένο γεγονός παρίσταται από το ζεύγος κατάστασης-ενέργειας (S_t, A_t).

Στην αμέσως επόμενη χρονική περίοδο t+1, στο άμεσο περιβάλλον εμφανίζεται μια κατάσταση $S_{t+1} \in S$. Αυτή τη στιγμή, ο πράκτορας δέχεται μια αριθμητική ανταμοιβή $R_{t+1} \in R$ ως αποτέλεσμα της ενέργειας A_t κατά τη κατάσταση S_t. Η διαδικασία της ανταμοιβής μπορεί να θεωρηθεί ότι ανταποκρίνεται στο αποτέλεσμα της εφαρμογής μιας αυθαίρετης συνάρτησης f που αντιστοιχεί τα ζεύγη κατάστασης-ενέργειας με μια ανταμοιβή (αποτέλεσμα). Σε κάθε χρονική περίοδο έχουμε επομένως:

$$f(S_t, A_t) = R_{t+1}$$

Το μοτίβο που περιγράφει τη βηματική διαδικασία της επιλογής μιας ενέργειας από μια δοσμένη κατάσταση, τη μετάβαση της σε μια νέα κατάσταση και την κατάκτηση μιας ανταμοιβής, μπορεί να περιγραφεί με την ακολουθία S_0, A_0, R_1, S_1, A_1, R_2, S_2, A_2, R_3, ….που εμφανίζεται στο ακόλουθο διάγραμμα



Σχήμα 1.1 Διάγραμμα Μαρκοβιανής Αλυσίδας

Το σχήμα 1.1 απεικονίζει κατά τη χρονική στιγμή t το άμεσο περιβάλλον του παρατηρητή βρίσκεται στη κατάσταση S_t, ο πράκτορας παρατηρεί την τρέχουσα κατάσταση και επιλέγει την ενέργεια A_t. Σαν αποτέλεσμα της ενέργειας η κατάσταση για το περιβάλλον μεταβάλλεται σε S_{t+1} και αποδίδει στον λαβών την απόφαση την ανταμοιβή R_{t+1}.

Η διαδικασία ξεκινάει και πάλι για το αμέσως επόμενο χρονικό διάστημα t+1, το οποίο πλέον δεν

είναι κάτι το μελλοντικό αλλά έχει γίνει πλέον το τρέχον διάστημα. Αμέσως μετά τη διακεκομμένη γραμμή, αριστερά του σχήματος 1.1, το t+1 μεταβάλλεται στο τρέχον χρονικό διάστημα t και τα S_{t+1} και R_{t+1} γίνονται αντίστοιχα S_t και R_t.

Καθώς τα σύνολα S και R είναι πεπερασμένα σύνολα, οι τυχαίες μεταβλητές S_t και R_t έχουν αυστηρά καθορισμένες κατανομές πιθανότητας, δηλαδή με άλλα λόγια όλες οι πιθανές τιμές έχουν και μια συγκεκριμένη πιθανότητα για να τους αποδοθούν. Αυτές οι κατανομές καθορίζονται μόνο από τις συγκεκριμένες τιμές του προηγούμενου ζεύγους κατάστασης και ενέργειας που υπήρχε κατά τη περίοδο t-1.

Αν υποθέσουμε ότι $s' \in S$ και $r \in R$, υπάρχει μια πιθανότητα να έχουμε $S_t = s'$ και $R_t = r$, και η πιθανότητα αυτή καθορίζεται, σύμφωνα με όσα αναφέρθηκαν πιο πάνω, από τις συγκεκριμένες τιμές της προηγούμενης κατάστασης $s \in S$ και ενέργειας $a \in A(s)$, όπου A(s) είναι το σύνολο των ενεργειών που προκύπτουν από την κατάσταση s. Η πιθανότητα μετάβασης στη κατάσταση s' με ανταμοιβή r λόγω της απόφασης (ενέργεια) α στη κατάσταση s, μπορεί να υπολογισθεί, μέσα από τις γνήσιες κατανομές, για όλες τις τιμές $s' \in S$, $s \in S$, $r \in R$ και $a \in A(s)$ ως:

$$p(s',r \mid s,a) = Pr\{S_t=s', R_t=r \mid S_{t-1} = s, A_{t-1} = a\}$$

2.4 Τα πλεονεκτήματα και οι αδυναμίες της ανάλυσης Markov

Η ανάλυση Markov έχει ένα σαφές πλεονέκτημα έναντι άλλων μεθόδων ανάλυσης (όπως η πολυεπίπεδη ανάλυση δεδομένων) όσον αφορά την ταχύτητα και την ακρίβεια, καθώς οι περισσότερες απαιτούν εκτελέσεις μεγαλύτερου αριθμού προσομοιώσεων, για να επιτευχθεί υψηλότερη ακρίβεια και, σε αντίθεση με την ανάλυση Markov, δεν παράγουν μια «ακριβή» απάντηση. Όπως και στην περίπτωση της εφαρμογής όλων των μεθόδων ανάλυσης, η ανάλυση Markov απαιτεί μεγάλη προσοχή κατά τη φάση κατασκευής του μοντέλου, καθώς η ακρίβεια του μοντέλου είναι πολύ σημαντική για την απόκτηση έγκυρων αποτελεσμάτων. Οι υποθέσεις που υπονοούνται στα μοντέλα Markov που σχετίζονται με την έλλειψη προγενέστερης μνήμης και κάποιες επιλογές, όπως η επιλογή της εκθετικής κατανομής για την αναπαράσταση των χρόνων μέχρι την αστοχία και την επισκευή, παρέχουν πρόσθετους περιορισμούς. Τα μοντέλα Markov μπορούν επομένως να γίνουν μη ρεαλιστικά εάν αυτές οι αρχικές υποθέσεις δεν αντικατοπτρίζουν επαρκώς τα χαρακτηριστικά ενός συστήματος και τον τρόπο λειτουργίας του στην πράξη. Ο κατασκευαστής ενός μοντέλου πρόβλεψης χρειάζεται επομένως να διαθέτει επαρκή εμπειρία προκειμένου να κερδίσει τα πλεονεκτήματα της ταχύτητας και της ακρίβειας που μπορεί να προσφέρει, η ανάλυση Markov. Επίσης, ενώ η ανάλυση Markov είναι μια ασφαλέστερη και πιο ευέλικτη προσέγγιση, δεν προσφέρει πάντα την ταχύτητα και την ακρίβεια που μπορεί να απαιτούνται σε συγκεκριμένες μελέτες συστήματος. Η ανάλυση Markov υποθέτει ότι η πιθανότητα μετάβασης σε μια συγκεκριμένη κατάσταση εξαρτάται μόνο από την προηγούμενη κατάσταση. Αυτό περιορίζει την ικανότητά της να αντιμετωπίσει περίπλοκες σχέσεις μεταξύ των καταστάσεων, καθώς δεν λαμβάνει υπόψη την πλήρη ιστορία των μεταβάσεων. Η ανάλυση Markov υποθέτει ότι οι πιθανότητες μετάβασης μεταξύ των καταστάσεων παραμένουν σταθερές στον χρόνο. Αυτή η υπόθεση μπορεί να μην ισχύει σε πραγματικά συστήματα, όπου οι συνθήκες μπορεί να μεταβάλλονται ειδικά στις στάσιμες αλυσίδες. Η ανάλυση Markov παραμένει ένα ισχυρό εργαλείο για τη μοντελοποίηση και την πρόβλεψη σε πολλούς τομείς, εφόσον χρησιμοποιείται με κατάλληλο τρόπο και με αντίληψη των περιορισμών της.

ΚΕΦΑΛΑΙΟ 3

Η Βελτιστοποίηση μέσω της Μαρκοβιανής Αλυσίδας

3.1 Κριτήριο Βελτιστοποίησης

Η επίλυση ενός Μαρκοβιανού Προβλήματος Απόφασης συνεπάγεται την αναζήτηση μιας πολιτικής. Ουσιαστικά είναι ένα σύνολο από κανόνες οι οποίοι καθορίζουν πώς θα πρέπει να επιλέγονται οι ενέργειες σε κάθε κατάσταση, ώστε να βελτιστοποιηθεί ένα κριτήριο βέλτιστης απόδοσης. Αυτό το κριτήριο στοχεύει στην αξιολόγηση της πολιτικής βασιζόμενο σε ένα μέτρο που θα παρέχει τις καλύτερες ακολουθίες των ανταμοιβών. Τυπικά, αντιστοιχεί στην αξιολόγηση μιας πολιτικής που βασίζεται στο αναμενόμενο άθροισμα των στιγμιαίων ανταμοιβών. Στην πραγματικότητα, η συσσώρευση των ανταμοιβών είναι συνήθως ένας τρόπος να μετρήσουμε την "καλή" απόδοση της πολιτικής σε τέτοιου είδους πρόβλημα. Τα κύρια κριτήρια που μελετώνται στη θεωρία των Μαρκοβιανών Διαδιασιών Αποφάσεων είναι

- πεπερασμένο κριτήριο: $E[r_0 + r_1 + r_2 + \cdots + r_{N-1} \mid s_0]$,
- (γ) κριτήριο αποπληθωριστικού κόστους: $E[r_0 + \gamma r_1 + \gamma^2 r_2 + \cdots + \gamma^t r_t + \cdots \mid s_0]$,
- κριτήριο με βάση το συνολικό μέσο κόστος: $E[r_0 + r_1 + r_2 + \cdots + r_t + \cdots \mid s_0]$,
- κριτήριο με βάση το μακροπρόθεσμο κόστος: $\lim_{n\to\infty}\frac{1}{n}E[r_0 + r_1 + r_2 + \cdots + r_{n-1} \mid s_0]$.

Τα ανωτέρω 4 κριτήρια μας επιτρέπουν να ορίσουμε την συνάρτηση αποτίμησης. Για μια δεδομένη πολιτική π και ένα κριτήριο μεταξύ των ανωτέρω, μια τέτοια συνάρτηση συνδέει οποιαδήποτε αρχική κατάσταση s με την τιμή του κριτηρίου που μας ενδιαφέρει να εξετάσουμε. Σε αντίθεση με τις ανταμοιβές που εκφράζουν την προσωρινή αξία μιας κατάστασης του περιβάλλοντος, οι τιμές εκφράζουν την μακροπρόθεσμη αξία μιας κατάστασης λαμβάνοντας υπόψη τις καταστάσεις που

ενδέχεται να προκύψουν στη συνέχεια και τις αντίστοιχες ανταμοιβές αθροιστικά (Sigaud & Baffet, 2010).

3.2 Συνάρτηση Απολαβής

Αυτή η συνάρτηση καθορίζει πώς θα υπολογιστεί η αναμενόμενη αξία ή ανταμοιβή για κάθε πολιτική. Η συνάρτηση απολαβής εξαρτάται από το συγκεκριμένο κριτήριο απόδοσης που χρησιμοποιείται. Για παράδειγμα, στο κριτήριο της μέγιστης αναμενόμενης ανταμοιβής, η συνάρτηση απολαβής θα πρέπει να υπολογίσει τον μέσο όρο των ανταμοιβών που προκύπτουν από την εκτέλεση της πολιτικής. Αν έχουμε μια πολιτική που αποφασίζει την επόμενη ενέργεια σε κάθε κατάσταση ενός Μαρκοβιανού Προβλήματος Απόφασης, μια συνάρτηση απολαβής μπορεί να λαμβάνει υπόψη την πιθανότητα μετάβασης σε κάθε νέα κατάσταση, την ανταμοιβή που συνδέεται με κάθε μετάβαση, και τον αποπληθωριστικό παράγοντα εάν χρησιμοποιείται κριτήριο που λαμβάνει υπόψη το μέλλον. Η συνάρτηση απολαβής είναι σημαντική γιατί επιτρέπει στον αλγόριθμο αναζήτησης πολιτικής να επιλέξει τη βέλτιστη πολιτική με βάση το κριτήριο απόδοσης που ορίζεται.

Έτσι σε μία πολιτική π αρχίζοντας από την κατάσταση s: π $V^{\pi} : S \rightarrow R$.

Ως V ορίζουμε το χώρο όλων των συναρτήσεων που αντιστοιχούν από το S στο R.

$$V^{\pi}(s) \leq V^{\pi^*}(s), \qquad \forall \pi \in \Pi, \forall s \in S$$

Αυτό που δηλώνει η έκφραση είναι ότι για κάθε πολιτική που υπάρχει στο σύνολο των πολιτικών Π και για κάθε κατάσταση s που υπάρχει στο σύνολο των καταστάσεων S, η τιμή της συνάρτησης απολαβής της πολιτικής πρέπει να είναι μικρότερη ή ίση με την τιμή της συνάρτησης αποτίμησης της βέλτιστης πολιτικής στην ίδια κατάσταση s (Sigaud & Baffet, (2010)).

3.2.1 Πεπερασμένο κριτήριο

Αυτό το κριτήριο αποσκοπεί στο να ελαχιστοποιήσει τον μέσο αριθμό βημάτων που απαιτούνται για να φτάσει το σύστημα σε μια τελική κατάσταση. Μια προηγούμενη υπόθεση που συνδέεται με αυτό το κριτήριο υποδηλώνει ότι ο λήπτης αποφάσεων πρέπει να ελέγχει το σύστημα N βημάτων, όπου το N είναι πεπερασμένο. Το πεπερασμένο κριτήριο ορίζει μια συνάρτηση τιμής που συνδέει κάθε κατάσταση s με το αναμενόμενο άθροισμα των επόμενων ανταμοιβών N που προκύπτει εφαρμόζοντας την πολιτική π από το s:

ΟΡΙΣΜΟΣ 1.2 (συνάρτηση απολαβής για το πεπερασμένο κριτήριο). Έστω $T = \{0, \ldots, N-1\}$, τότε γράφουμε

$$\forall s \in S, V_N^{\pi}(s) = E_{\pi}\left[\sum_{t=0}^{N-1} r_t \mid s_0 = s\right]$$

Σε αυτόν τον ορισμό, το $E_{\pi}[\cdot]$ υποδηλώνει το σύνολο των πιθανών αποτελεσμάτων του διανύσματος MDP, ακολουθώντας την πολιτική π. Το E_{π} σχετίζεται με την κατανομή πιθανοτήτων P_{π} σε αυτά τα αποτελέσματα.

Στην πράξη, κάποιες φορές, μπορεί να είναι ωφέλιμο να προσθέσουμε μια τελική ανταμοιβή r_N στο κριτήριο. Αυτή η ανταμοιβή είναι μόνο συνάρτηση της τελικής κατάστασης S_N. Για το σκοπό αυτό, θεωρούμε ένα τεχνητό πρόσθετο βήμα όπου $\forall(s, a) \in S \times A$, $r_N(s, a) = r_N(s)$. Αυτό το είδος της απαίτησης προκύπτει ιδιαίτερα όταν ο στόχος είναι να δοθεί εντολή σε ένα σύστημα, προς μια κατάσταση στόχου, σε N βήματα και με βέλτιστο (λιγότερο) κόστος (Sigaud & Baffet, 2010).

Αυτό το κριτήριο μπορεί να είναι χρήσιμο σε πρακτικές εφαρμογές όπου ο στόχος είναι να εκτελεστεί μια εργασία σε όσο το δυνατόν λιγότερο χρόνο ή σε περιπτώσεις όπου ο χρόνος είναι κρίσιμος για την επιτυχή ολοκλήρωση μιας διαδικασίας (Sigaud & Baffet, 2010)).

3.2.2 Κριτήριο αποπληθωριστικού κόστους

Ο αποπληθωριστικός παράγοντας επηρεάζει τη διαδικασία λήψης αποφάσεων στις διαδικασίες Μαρκοβιανών αποφάσεων και παίζει κρίσιμο ρόλο στον υπολογισμό της αξίας μιας κατάστασης ή της αξίας μιας πολιτικής. Συμβολίζεται με το γράμμα γ και παίρνει τιμές από 0 έως 1. Όταν το γ = 0, δεν δίνεται καμία βαρύτητα στις μελλοντικές ανταμοιβές, ενώ όταν γ = 1, όλες οι ανταμοιβές έχουν την ίδια σημασία, ανεξάρτητα από το πόσο μακριά βρίσκονται στο μέλλον. Συνδέεται κυρίως με τη συνάρτηση απολαβής δοθείσης της πολιτικής π με το αναμενόμενο άθροισμα των ανταμοιβών όπου κάθε ανταμοιβή αντιστοιχεί σε ένα βάρος από έναν αποπληθωριστικό παράγοντα. Το βασικό πρόβλημα σε μία διαδικασία Μαρκοβιανών αποφάσεων, έγκειται στην εύρεση μίας πολιτικής π από τον λήπτη αποφάσεων. Η συνάρτηση προσδιορίζει την ενέργεια που πρέπει να παρθεί από τον λήπτη αποφάσεων, στην κατάσταση s. Στόχος είναι η επιλογή της πολιτικής η οποία θα μεγιστοποιήσει το αναμενόμενο άθροισμα των επιβραβεύσεων, σε άπειρο ή περιορισμένο χρόνο. Χρησιμοποιείται, επίσης, και ένας αποπληθωριστικός παράγοντας, ώστε να δείξει στον υπεύθυνο λήψης αποφάσεων σε τι βαθμό τον ενδιαφέρει η επιβράβευση την τωρινή χρονική στιγμή, σε σχέση με μία μελλοντική.

Ορισμός 1.3 (συνάρτηση απολαβής για τον αποπληθωριστικό παράγοντα).
Έστω $0 \leq \gamma \leq 1$, έχουμε ότι

$$\forall s \in S,\ V_{\gamma}^{\pi}(s) = E^{\pi}\left[\sum_{t=0}^{\infty} \gamma^{t} r_{t} \mid s_{0} = s\right],\ \forall s \in S$$

Ο παράγοντας γ αντιπροσωπεύει την τιμή, που λαμβάνεται υπόψη από τη στιγμή t, μιας μοναδιαίας ανταμοιβής τη στιγμή t + 1. Εάν λαμβανόταν τη στιγμή t + τ, αυτή η ίδια ανταμοιβή που ισούται με ένα λαμβάνονταν ως γ_{τ} . Ως εκ τούτου, αυτό ευνοεί τις ανταμοιβές που λαμβάνονται νωρίς στη διαδικασία (Sigaud & Baffet, 2010)).

3.2.3 Κριτήριο με βάση το συνολικό μέσο κόστος

Όπως αναφέρεται στους Sigaud & Baffet (2010) προκειμένου να γίνει σωστή αξιολόγηση της πολιτικής αυτό το κριτήριο αποτελείται από το άθροισμα των ανταμοιβών που λαμβάνονται, με την προϋπόθεση ότι οι ανταμοιβές αυτές έχουν κατάλληλα αποπληθωριστεί για να ληφθεί υπόψη η αβεβαιότητα και η προτίμηση για αμέσως επόμενες ανταμοιβές έναντι μελλοντικών. Δίνοντας την τιμή γ=1 μπορούμε να εξετάσουμε την ακόλουθη σχέση.

Ορισμός 1.4

$$V_{\pi}(s)= E^{\pi}[\sum_{t=0}^{\infty} r_t \ |s_0 = s]$$

Το συγκεκριμένο κριτήριο χρησιμοποιείται σε πεπερασμένο χρόνο χωρίς όμως να γνωρίζουμε την ακριβή τιμή του. Σ' αυτή την περίπτωση η διαδικασία τερματίζει μετά από περασμένο χρόνο με πιθανότητα ίσο με ένα (Sigaud & Baffet, 2010).

3.2.4 Κριτήριο με βάση το μακροπρόθεσμο μέσο κόστος

Όταν οι αποφάσεις οι οποίες καλούμαστε να λάβουμε με τον αποπληθωριστικό παράγοντα κοντά στην τιμή ένα ή όταν αδυνατούμε να αξιολογήσουμε ακριβώς τις ανταμοιβές τότε είναι προτιμότερο να υπολογίσουμε ένα κριτήριο το οποίο να ανταποκρίνεται στη μέση τιμή των ληφθέντων ανταμοιβών αντί να υπολογίζουμε το άθροισμα των βαρών. Έτσι ορίζεται το κριτήριο της μέσης τιμής ως $\rho_{\pi}(s)$ όπου μπορούμε να σχηματίσουμε μια συνάρτηση τιμής συνδέοντας καταστάσεις στο μέσο κέρδος ανά βήμα, σε άπειρο χρονικό ορίζοντα αν εφαρμόσουμε την πολιτική π.

Ορισμός 1.5 (Συνάρτηση απολαβής αναμενόμενης τιμής)

$$\rho^{\pi}(s) = \lim_{n\to\infty} E^{\pi}\left[\frac{1}{n}\sum_{t=0}^{n-1} r_t | s_0 = s\right]$$

Μια πολιτική π∗ λέγεται ότι έχει βέλτιστο κέρδος για το μέσο κριτήριο εάν $\rho^{\pi*}(s) \geq \rho^{\pi}(s)$ για οποιαδήποτε πολιτική π και κατάσταση s (Sigaud & Baffet, 2010).

3.3 Αλγόριθμοι Βελτιστοποίησης για Μαρκοβιανές Διαδικασίες Αποφάσεων

3.3.1 Αλγόριθμος Πεπερασμένου Χρονικού Ορίζοντα

Ο αλγόριθμος πεπερασμένου κριτηρίου υπολογίζει αναδρομικά τις βέλτιστες συναρτήσεις των τιμών $V_1^*,\ldots V_N^*$ από το τελευταίο βήμα έως το πρώτο όπως περιγράφεται ακολούθως η διαδικασία:

$V_0 \leftarrow\!\!- 0$

για $n \leftarrow\!\!- 0$ με N-1 ΕΚΤΕΛΕΣΕ

για $s \in S$ ΕΚΤΕΛΕΣΕ

$$V_{n+1}^* = \max_{a \in A}\left\{ r_{N-1-n}(s,\alpha) + \sum_{S'} p_{N-1-n}(s'|s,\alpha) V_n^*(s') \right\}$$

$$\pi_{N-1-n}(s) \in \arg\max_{a \in A}\left\{ r_{N-1-n}(s,a) + \sum_{s'} p_{N-1-\nu}(s'|s,a) V_n^*(s') \right\}$$

ΕΠΑΝΑΛΗΨΗ V^*, π^*

Τα βήματα του αλγορίθμου αρχίζουν με τον καθορισμό των μεταβλητών απόφασης, των περιορισμών και της συνάρτησης κόστους που πρέπει να βελτιστοποιηθεί. Στη συνέχεια αρχικοποίηση της συνάρτησης αξίας ή κόστους για την τελική χρονική περίοδο. Αυτό είναι συνήθως απλό, καθώς στην τελευταία περίοδο δεν υπάρχουν μελλοντικές αποφάσεις προς λήψη. Έπειτα Επανάληψη αναδρομικά προς τα πίσω στον χρόνο, από την τελευταία περίοδο έως την αρχική, επιλύοντας τις βέλτιστες αποφάσεις και ενημερώνοντας τη συνάρτηση αξίας ή κόστους σε κάθε βήμα .Όταν η επανάληψη φτάσει στην αρχική περίοδο, παίρνουμε τη βέλτιστη απόφαση και τη συνάρτηση αξίας ή κόστους. Τερματίζει όταν η αναδρομή φτάσει στην αρχική περίοδο, παίρνουμε τη βέλτιστη απόφαση και τη συνάρτηση αξίας ή κόστους. (Sigaud & Baffet, 2010)

3.3.2 Αλγόριθμος Αποπληθωρισμένου Κόστους γ

Ο αλγόριθμος αποπληθωρισμένου κόστους χρησιμοποιείται για την εύρεση της βέλτιστης πολιτικής λήψης αποφάσεων σε προβλήματα Μαρκοβιανών Διαδικασιών Αποφάσεων (Markov Decision Processes), όπου εφαρμόζεται ο αποπληθωριστικός παράγοντας στις μελλοντικές ανταμοιβές όπως περιγράφεται ακολούθως η διαδικασία:

ΕΙΣΗΓΑΓΕ $\pi_0 \in D$

$n \leftarrow\!\!- 0$

ΕΠΑΝΕΛΑΒΕ

ΕΚΤΕΛΕΣΕ

$$V_n = r(s,\pi_n(s)) + \gamma \sum_{s \in S} p(s'|\pi_n(s)) V_n(s'), \qquad \forall s \in S$$

ΓΙΑ $s \in S$ ΕΚΤΕΛΕΣΕ

$$\pi_{n+1}(s) \in \arg\max_{a \in A} \left\{ r(s,a) + \gamma \sum_{s' \in S} p(s'|s,a) V_n(s') \right\}$$

$n \leftarrow\!\!- n+1$

ΕΩΣ $\pi_n = \pi_{n+1}$

ΕΠΑΝΑΛΗΨΗ V_n, π_{n+1}

(Sigaud & Baffet, 2010))

Τα βήματα του αλγορίθμου αρχίζουν με μια αρχική πολιτική, συνήθως επιλεγμένη τυχαία ή βασισμένη σε προηγούμενη γνώση. Γίνεται η αξιολόγηση της συνάρτησης αξίας για την τρέχουσα πολιτική. Αυτή περιλαμβάνει την επίλυση της εξίσωσης του Bellman επαναληπτικά έως ότου η νέα πολιτική να είναι όμοια με την προηγούμενη. Η εξίσωση του Bellman για τον αποπληθωριστικό παράγοντα είναι:

$$V(s) = max_a(R(s,a) + \gamma \sum_{s'} P(s' | s,a) V(s'))$$

Όπου:

- $V^{\pi}(s)$είναι η τιμή της κατάστασης s υπό την πολιτική π.
- $\pi(a|s)$είναι η πιθανότητα επιλογής της ενέργειας a στην κατάσταση s σύμφωνα με την

πολιτική π.

- $P(s'|s,a)$ είναι η πιθανότητα μετάβασης στην κατάσταση s' όταν επιλέγεται η ενέργεια a από την κατάσταση s.

$R(s,a,s')$ είναι η άμεση ανταμοιβή που λαμβάνεται μετά τη μετάβαση από την κατάσταση s στην κατάσταση s' παίρνοντας την ενέργεια a.

γ είναι ο αποπληθωριστικός παράγοντας ($0 < \gamma < 1$) που αντιπροσωπεύει τη σημασία των μελλοντικών ανταμοιβών σε σχέση με τις άμεσες ανταμοιβές.

Ο καθορισμός της πολιτικής γίνεται επιλέγοντας την ενέργεια σε κάθε κατάσταση που μεγιστοποιεί την αναμενόμενη ανταμοιβή. Η νέα πολιτική καθορίζεται ως εξής:

$$\pi'(s)=\text{argmax}_a \sum_{s'} P(s'|s,a)[R(s,a,s')+\gamma V^{\pi}(s')]$$

Ελέγχεται αν η νέα πολιτική είναι ίδια με την προηγούμενη πολιτική. Αν οι πολιτικές είναι ίδιες, ο αλγόριθμος έχει συγκλίνει και η τρέχουσα πολιτική είναι βέλτιστη. Διαφορετικά, επιστρέφει στο βήμα όπου εκτελείται ξανά η αξιολόγηση της πολιτικής (Sigaud & Baffet, 2010)).

3.3.3 Αλγόριθμος βελτίωσης των πολιτικών-αποπληθωρισμένο κριτήριο

Ο αλγόριθμος με κριτήριο το συνολικό μέσο κόστος λειτουργεί παρόμοια με τον αλγόριθμο με κριτήριο τον αποπληθωριστικό παράγοντα, αλλά αντί για μείωση στις μελλοντικές ανταμοιβές, επιδιώκει να μεγιστοποιήσει ή να ελαχιστοποιήσει τη συνολική απολαβή σε κάθε πολιτική. Αυτό σημαίνει ότι η αξία κάθε κατάστασης δεν θα "μειώνεται" ανάλογα με το πόσο μακριά βρίσκεται στο μέλλον, αλλά θα λαμβάνει υπόψη τη συνολική απολαβή που αναμένεται από αυτή την κατάσταση όπως περιγράφεται ακολούθως η διαδικασία:

ΕΙΣΗΓΑΓΕ $\pi_0 \in D$ με $r_{\pi 0} \geq 0$

$n \leftarrow\!\!- 0$

ΕΠΑΝΕΛΑΒΕ

ΥΠΟΛΟΓΙΣΕ

$$V_n = r(s,\pi_n(s)) + \sum_{s' \in S} p(s'|\pi_n(s)) V_n(s'), \qquad s \in S$$

ΓΙΑ $s \in S$ ΕΚΤΕΛΕΣΕ

$$\pi_{n+1}(s) \in \arg\max_{a \in A} \left\{ r(s,a) + \sum_{s' \in S} p(s'|s,a) V_n(s') \right\}$$

$$\pi_{n+1}(s) = \pi_n(s)$$

$n \leftarrow\!\!- n+1$

ΕΩΣ $\pi_n = \pi_{n+1}$

ΕΠΑΝΑΛΗΨΗ V_n, π_{n+1}

(Sigaud & Baffet, 2010)

Παρατηρούμε ότι αρχικοποιεί την αρχική πολιτική και τις τιμές των καταστάσεων και υπολογίζει την αξία της κάθε κατάστασης για την τρέχουσα πολιτική, λύνοντας την εξίσωση Bellman για την συνολική απολαβή. Αυτή η εκτίμηση μπορεί να επαναληφθεί μέχρι να συγκλίνει η τιμή κάθε κατάστασης. Έπειτα ενημερώνεται η πολιτική επιλέγοντας την ενέργεια που μεγιστοποιεί την συνολική απολαβή σε κάθε κατάσταση.

Αυτή η ενημέρωση γίνεται με βάση την εκτίμηση των καταστάσεων από το προηγούμενο βήμα. Εάν η νέα πολιτική είναι ίδια με την προηγούμενη, ο αλγόριθμος τερματίζει, καθώς έχει συγκλίνει στη βέλτιστη πολιτική. Διαφορετικά, επιστρέφει στο βήμα όπου εκτελείται η εκτίμηση της πολιτικής

π. Αυτός ο αλγόριθμος συνεχίζει να εκτελείται μέχρι να συγκλίνει στη βέλτιστη πολιτική, που μεγιστοποιεί την συνολική ανταμοιβή στο πλαίσιο των θετικών αποφάσεων (Sigaud & Baffet, 2010).

3.3.4 Τροποποιημένος Αλγόριθμος βελτίωσης πολιτικής

Ο αλγόριθμος με βάση το μακροπρόθεσμο μέσο κόστος λειτουργεί διαφορετικά από τους κλασικούς αλγορίθμους πολιτικής επανάληψης που αναφέρθηκαν προηγουμένως. Στα κριτήρια, όπως η συνολική απολαβή ή η τιμή του αποπληθωριστικού παράγοντα, ή η αξία των καταστάσεων αναφέρεται στο σύνολο των ανταμοιβών που λαμβάνονται. Ωστόσο, στο κριτήριο της αναμενόμενης τιμής, ενδιαφερόμαστε για τη μέση ανταμοιβή που λαμβάνεται σε κάθε βήμα της εκτέλεσης της πολιτικής, ανεξάρτητα από το συνολικό ποσό ανταμοιβής που θα ληφθεί κατά τη διάρκεια όλης της διαδικασίας. Ο αλγόριθμος για το κριτήριο της μέσης ανταμοιβής που περιγράφεται ακολούθως περιλαμβάνει συνήθως τα ακόλουθα βήματα:

ΕΙΣΗΓΑΓΕ $V_0 \in V$

ΣΥΝΘΗΚΗ ΨΕΥΔΗΣ

Για $n \leftarrow\!\!- 0$

ΕΠΑΝΕΛΑΒΕ

ΕΚΤΕΛΕΣΕ ΓΙΑ $s \in S$

$$\pi_{n+1}(s) \in \arg\max_{a \in A} \left\{ r(s,a) + \gamma \sum_{s' \in S} p(s'|s,a) V_n(s') \right\}$$

$$\pi_{n+1}(s) = \pi_n(s) \; \alpha\nu \; \epsilon\iota\nu\alpha\iota \; \delta\upsilon\nu\alpha\tau\acute{o}$$

$$V_n^0 = \max_{a \in A} \left\{ r(s,a) + \sum_{s' \in S} p(s'|s,a) V_n(s') \right\}$$

ΑΝ ΕΥΡΟΣ $(V_n^0 - V_n) < \epsilon$ ΤΟΤΕ ΣΥΝΘΗΚΗ ΑΛΗΘΗΣ

ΑΛΛΙΩΣ $m \leftarrow\!\!- 0$

ΕΠΑΝΕΛΑΒΕ

ΕΚΤΕΛΕΣΕ ΓΙΑ $s \in S$

$$V_n^{m+1} = r(s, \pi_{n+1}(s)) + \sum_{s' \in S} p(s'|\pi_{n+1}(s)) V_n^m(s')$$

$m \leftarrow\!\!- m+1$

ΕΩΣ ΕΥΡΟΣ $(V_n^{m+1} - V_n^m) < \delta$

$V_n \leftarrow\!\!- V_n^m$

$n \leftarrow\!\!- n+1$

ΕΩΣ ΣΥΝΘΗΚΗ ΑΛΗΘΗΣ

ΕΠΕΣΤΡΕΨΕ V_n, π_{n+1}

(Sigaud & Baffet, 2010)

Αρχικοποιεί την αρχική πολιτική και τις τιμές των καταστάσεων. Υπολογίζει την μέση ανταμοιβή που λαμβάνεται από κάθε κατάσταση κατά την εκτέλεση της τρέχουσας πολιτικής.

Αυτή η εκτίμηση μπορεί να επαναληφθεί μέχρι να συγκλίνει η τιμή της μέσης ανταμοιβής για κάθε κατάσταση. Ενημερώνει την πολιτική επιλέγοντας την ενέργεια που μεγιστοποιεί τη μέση ανταμοιβή σε κάθε κατάσταση. Εάν η νέα πολιτική είναι ίδια με την προηγούμενη, ο αλγόριθμος τερματίζει, καθώς έχει συγκλίνει στη βέλτιστη πολιτική. Διαφορετικά, επιστρέφει στο βήμα όπου γίνεται η εκτίμηση της πολιτικής. Αυτός ο αλγόριθμος συνεχίζει να εκτελείται μέχρι να συγκλίνει στη βέλτιστη πολιτική που μεγιστοποιεί τη μέση ανταμοιβή κατά τη διάρκεια της εκτέλεσης της πολιτικής (Sigaud & Baffet, 2010).

Παράδειγμα 3.1

Ένα παράδειγμα για το πώς λειτουργεί συγκεκριμένα ο αλγόριθμος με βάση τον αλγόριθμο του μακροπρόθεσμου μέσου κόστους που προαναφέρθηκε αφορά τη συντήρηση ενός αυτοκινήτου του (Howard, 1960). Το πρόβλημα αντικατάστασης ή πρόβλημα βελτιστοποίησης-συντήρησης, είναι πράγματι πολύ πρακτικό για διάφορους κλάδους, ιδιαίτερα αυτούς που βασίζονται σε μεγάλο βαθμό στον βιομηχανικό εξοπλισμό. Αυτό το πρόβλημα εγείρει το ερώτημα αν έχουμε αυτή τη στιγμή στην κατοχή μας το αυτοκίνητο συγκεκριμένων ετών, θα το αντικαταστήσουμε ή θα το κρατήσουμε και να το συντηρήσουμε αναλαμβάνοντας τη φθορά με την πάροδο του χρόνου λόγω χρήσης ή άλλων παραγόντων και να ληφθεί στη συνέχεια η απόφαση. Άρα το κύριο ερώτημα στο πρόβλημα είναι εάν θα συνεχιστεί η χρήση του υπάρχοντος αυτοκινήτου ή η αντικατάστασή του με νέο. Υποτίθεται ότι το αυτοκίνητο είναι 10 ετών. Κάθε 3 μήνες με την δεδομένη τρέχουσα κατάσταση θα πρέπει να λάβουμε την απόφαση αν θα το κρατήσουμε ή θα το πουλήσουμε για τη δεδομένη στιγμή. Για να επιλύσει ο αλγόριθμος το πρόβλημα αντικατάστασης, πρέπει να μοντελοποιηθεί ως Διαδικασία Μαρκοβιανής Απόφασης όπου: Οι Καταστάσεις αντιπροσωπεύουν την ηλικία του αυτοκινήτου σε περιόδους των τριών μηνών οι οποίες λαμβάνουν τιμή από 0 έως 40. Διευκρινίζεται ότι οι καταστάσεις στο πρόβλημα θα είναι πεπερασμένες όταν φτάσει στην ηλικία των 40.

Ο χώρος καταστάσεων γίνεται ως εξής:

C_i, είναι η τιμή αγοράς αυτοκινήτου στην ηλικία i,

T_i, η τρέχουσα τιμή πώλησης του αυτοκινήτου στην ηλικία i,

E_i, τα λειτουργικά έξοδα στην ηλικία i μέχρι να φτάσει στην ηλικία i+1

p_i, η πιθανότητα επιβίωσης του αυτοκινήτου στην ηλικία i+1 χωρίς να απαιτείται κάποιο απαγορευμένο κόστος επισκευής

Ορίζουμε την πιθανότητα να περιοριστεί ο αριθμός των καταστάσεων. Άρα η πιθανότητα$p_{40}=0$ αφού μια πιθανή βλάβη του αυτοκινήτου πάει κατευθείαν στην κατάσταση 40. Έστω ότι $\kappa=1$ το αυτοκίνητο θα είναι στην κατοχή για ένα ακόμα τρίμηνο και εναλλακτικά $\kappa>1$ πρόκειται να αγοραστεί αυτοκίνητο ηλικίας κ-2 ηλικίας 39 μηνών. Η μέγιστη χρονολογία υπολογίζεται στα 10 έτη (40 3μηνα) όπως ο πίνακας 4.

Οι ενέργειες αντιπροσωπεύουν τις αποφάσεις διατήρησης του αυτοκινήτου ή αντικατάστασής του με νέο. Οι μεταβάσεις αντιπροσωπεύουν τις πιθανότητες μετάβασης του αυτοκινήτου από τη μια κατάσταση στην άλλη με την πάροδο του χρόνου λόγω φθοράς. Οι ανταμοιβές αντιπροσωπεύουν το κόστος που σχετίζεται με τη διατήρηση του αυτοκινήτου, την αντικατάστασή του ή άλλους σχετικούς παράγοντες, όπως το κόστος παύσης λειτουργίας. Η μέθοδος βελτίωσης της πολιτικής μπορεί στη συνέχεια να χρησιμοποιηθεί για να βελτιώσει επαναληπτικά την πολιτική λήψης αποφάσεων σχετικά με το πότε πρέπει να αντικατασταθεί το αυτοκίνητο.

Αυτό περιλαμβάνει την εναλλαγή μεταξύ αξιολόγησης πολιτικής, όπου η συνάρτηση τιμής υπολογίζεται για μια δεδομένη πολιτική, και βελτίωσης πολιτικής, όπου η πολιτική ενημερώνεται με βάση τη συνάρτηση απολαβής. Κατά την αξιολόγηση πολιτικής, εκτιμάται το αναμενόμενο συνολικό κόστος ή ανταμοιβή που σχετίζεται με κάθε ζεύγος κατάστασης-ενέργειας. Στη βελτίωση της πολιτικής, η πολιτική λήψης αποφάσεων ενημερώνεται για να επιλεχτούν ενέργειες που ελαχιστοποιούν το κόστος ή μεγιστοποιούν τις ανταμοιβές με βάση τις εκτιμώμενες τιμές. Εφαρμόζοντας επαναληπτικά βήματα αξιολόγησης και βελτίωσης πολιτικής, η μέθοδος επανάληψης πολιτικής συγκλίνει σε μια βέλτιστη πολιτική που καθορίζει την καλύτερη πορεία ενέργειας για κάθε κατάσταση, δηλαδή εάν θα διατηρήσει το αυτοκίνητο ή θα αντικατασταθεί και εάν αντικατασταθεί, ποια χαρακτηριστικά θα πρέπει να έχει το νέο αυτοκίνητο.

Συνολικά, η μέθοδος επανάληψης πολιτικής παρέχει μια συστηματική προσέγγιση για την επίλυση του προβλήματος αντικατάστασης βελτιστοποιώντας τη διαδικασία λήψης αποφάσεων σχετικά με

τη συντήρηση και την αντικατάσταση εξοπλισμού, οδηγώντας τελικά σε πιο αποτελεσματική χρήση πόρων και εξοικονόμηση κόστους για τις βιομηχανίες. Συνολικά, η χρήση της Διαδικασίας Μαρκοβιανής Απόφασης για τη συντήρηση ενός αυτοκινήτου επιτρέπει τη λήψη αποφάσεων βάσει του μακροπρόθεσμου στόχου της ελαχιστοποίησης του κόστους συντήρησης, λαμβάνοντας υπόψη τις πιθανές ανάγκες συντήρησης και το κόστος των ενεργειών.

Συμπερασματικά, κάποιος πρέπει να κρατήσει το αυτοκίνητο του αν είναι σε ηλικία μισού έτους και πάνω ή είναι 6 και μισό χρονών. Να ανταλλάξει το αυτοκίνητό που βρίσκεται στην κατοχή του αυτή τη στιγμή για ένα αυτοκίνητο που είναι τρία χρονών. Αυτός είναι ένας κοινός τρόπος να αποκτήσει ένα νέο αυτοκίνητο χωρίς να πληρώσει κάποιος το πλήρες κόστος ενός ολοκαίνουργιου οχήματος. Η αξία του αυτοκινήτου φαίνεται στον πίνακα 3.

ΠΙΝΑΚΑΣ 3

Δεδομένα Συντήρησης Αυτοκινήτου									
Ηλικία ανά περίοδο	Κόστος	Εμπορεύσιμη αξία	Λειτουργικά έξοδα	Πιθανότητα Επιβίωσης	Ηλικία ανά περίοδο	Κόστος	Εμπορεύσιμη αξία	Λειτουργικά έξοδα	Πιθανότητα Επιβίωσης
0	2000	1600	50	1					
1	1840	1460	53	0,999	21	345	240	115	0,925
2	1680	1340	56	0,998	22	330	225	118	0,919
3	1560	1230	59	0,997	23	315	210	121	0,910
4	1300	1050	62	0,996	24	300	200	125	0,900
5	1220	980	65	0,994	25	290	190	129	0,890
6	1150	910	68	0,991	26	280	180	133	0,880
7	1080	840	71	0,988	27	265	170	137	0,865
8	900	710	75	0,985	28	250	160	141	0,850
9	840	650	78	0,983	29	240	150	145	0,820
10	780	600	81	0,980	30	230	145	150	0,790
11	730	550	84	0,975	31	220	140	155	0,760
12	600	480	87	0,970	32	210	135	160	0,730
13	560	430	90	0,965	33	200	130	167	0,660
14	520	390	93	0,960	34	190	120	175	0,590
15	480	360	96	0,955	35	180	115	182	0,510
16	440	330	100	0,950	36	170	110	190	0,430
17	420	310	103	0,945	37	160	105	205	0,300
18	400	290	106	0,940	38	150	95	220	0,200
19	380	270	109	0,935	39	140	87	235	0,100
20	360	255	112	0,930	40	130	80	250	0,000

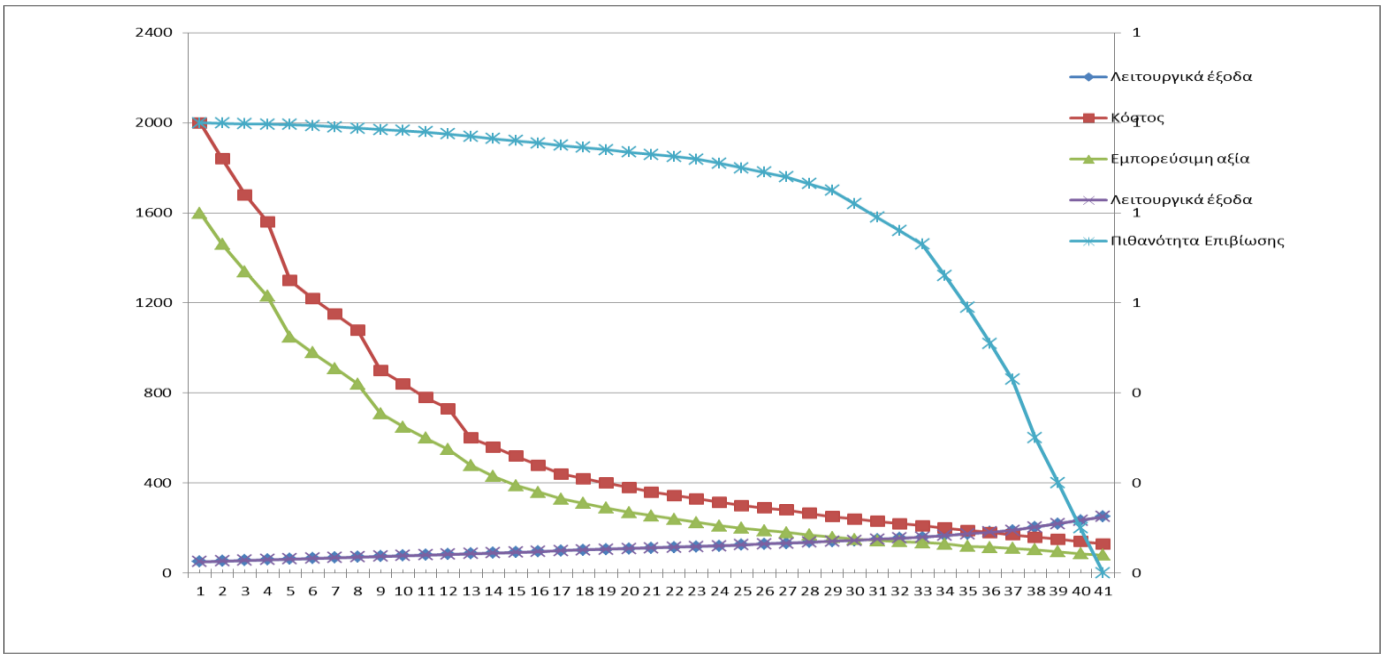
(Howard, 1960)

Ο πίνακας 3 σύμφωνα με τον Ronald A. Howard (1960) παρουσιάζει την εξέλιξη του κόστους, της εμπορεύσιμης αξίας, των λειτουργικών εξόδων και της πιθανότητας επιβίωσης των αυτοκινήτων σε

βάθος 10 ετών (1 περίοδος/3μήνες). Παρατηρούμε ότι με αυτά τα δεδομένα καλείται κάποιος να λάβει απόφαση σχετικά με την αντικατάσταση ή τη συντήρηση των αυτοκινήτων.

- Κόστος και Εμπορεύσιμη Αξία: Με την πάροδο του χρόνου, παρατηρείται μια μείωση στην εμπορεύσιμη αξία των αυτοκινήτων, ενώ το κόστος μειώνεται σε μικρότερο βαθμό.
- Λειτουργικά Έξοδα: Τα λειτουργικά έξοδα αυξάνονται σταδιακά με την πάροδο του χρόνου, γεγονός που επηρεάζει την βιωσιμότητα και της διατήρησης των παλαιότερων αυτοκινήτων.
- Πιθανότητα Επιβίωσης: Η πιθανότητα επιβίωσης μειώνεται με την ηλικία του αυτοκινήτου, υποδεικνύοντας την αυξανόμενη πιθανότητα μη ομαλής λειτουργίας. Έτσι καταλήγουμε στο συμπέρασμα ότι βάσει των δεδομένων, είναι σαφές ότι η αντικατάσταση του αυτοκινήτου πριν από το 5ο έτος μπορεί να είναι πιο οικονομική λόγω της μειωμένης εμπορεύσιμης αξίας και των αυξημένων λειτουργικών εξόδων μετά από αυτό το σημείο. Η στρατηγική συντήρησης θα πρέπει να λαμβάνει υπόψη την αυξανόμενη πιθανότητα επιβίωσης και τα λειτουργικά έξοδα, προκειμένου να μεγιστοποιηθεί η απόδοση της επένδυσης στο αυτοκίνητο. Μέσω του δυναμικού προγραμματισμού που ενσωματώνει αυτά τα δεδομένα θα μπορούσε να προσφέρει μια βέλτιστη πολιτική αντικατάστασης, μειώνοντας τα συνολικά έξοδα και αυξάνοντας την αποδοτικότητα. Η ανάλυση του πίνακα παρέχει χρήσιμες πληροφορίες για τη λήψη στρατηγικών αποφάσεων σχετικά με τη συντήρηση και αντικατάσταση των αυτοκινήτων, εξοικονομώντας πόρους και βελτιώνοντας την αποδοτικότητα.

ΔΙΑΓΡΑΜΜΑ 1



(Ronald A. Howard, 1960)

Μέσω του διαγράμματος 1 γίνεται η οπτικοποιήση των πινάκων 3 & 4 όπου παρατηρείται η έξέλιξη του αυτοκινήτου κατά τη διάρκεια των 40 περιόδων. Συνολικά, το διάγραμμα φαίνεται να παρουσιάζει μια κατάσταση όπου αρχικά τα κόστη και οι εμπορεύσιμες αξίες μειώνονται, ενώ τα λειτουργικά έξοδα παραμένουν σταθερά. Η πιθανότητα επιβίωσης παραμένει υψηλή μέχρι το τέλος, όπου μειώνεται απότομα. Αυτό δείχνει μια σταδιακή μείωση της αξίας και του κόστους, ενώ ο ιδιοκτήτης προσπαθεί να διατηρήσει τα λειτουργικά έξοδα σταθερά, αλλά τελικά αντιμετωπίζει μια κρίσιμη πτώση στην πιθανότητα επιβίωσης.

· Λειτουργικά έξοδα: Αυτή η καμπύλη παραμένει σταθερή, δείχνοντας ότι τα λειτουργικά έξοδα δεν αλλάζουν με την πάροδο του χρόνου.

· Κόστος : Αυτή η καμπύλη ξεκινάει από μια υψηλή τιμή και μειώνεται με την πάροδο του χρόνου, υποδεικνύοντας ότι το κόστος μειώνεται καθώς περνούν οι περίοδοι.

· Εμπορεύσιμη αξία : Αυτή η καμπύλη επίσης μειώνεται σταδιακά, δείχνοντας ότι η αξία του προϊόντος ή της επιχείρησης μειώνεται με τον χρόνο.

· Λειτουργικά έξοδα 1 : Αυτή η καμπύλη είναι επίσης σταθερή και παραμένει χαμηλότερη από τα Λειτουργικά Έξοδα.

· Πιθανότητα Επιβίωσης : Αυτή η καμπύλη παραμένει σταθερή και υψηλή για τις περισσότερες περιόδους και μειώνεται απότομα στο τέλος, υποδεικνύοντας μια ξαφνική πτώση στην πιθανότητα επιβίωσης στο τέλος της περιόδου.

ΠΙΝΑΚΑΣ 4

Αποτελέσματα Αντικατάστασης Αυτοκινήτου															
	Επανάληψη 1		Επανάληψη 2		Επανάληψη 3		Επανάληψη 4		Επανάληψη 5		Επανάληψη 6		Επανάληψη 7		
	Κέρδος	250	Κέρδος	193,89	Κέρδος	162,44	Κέρδος	157,07	Κέρδος	151,05	Κέρδος	150,99	Κέρδος	150,95	Προσαρμοσμένη Αξία
Κατάσταση	Απόφαση	Αξία	Απόφαση	Αξία	Απόφαση	Αξία	Απόφαση	Αξία	Απόφαση	Αξία	Απόφαση	Αξία	Απόφαση	Αξία	
1	36	1374	20	1380	19	1380	12	1380	12	1380	12	1380	12	1380	1460
2	36	1254	20	1260	19	1260	12	1260	12	1260	12	1260	12	1260	1340
3	36	1144	20	1150	19	1150	12	1150	12	1150	12	1150	K	1161	1241
4	36	964	20	970	K	1037	12	970	K	1003	K	1072	K	1072	1152
5	36	894	20	900	K	940	12	900	K	917	K	987	K	987	1067
6	36	824	20	830	K	848	12	830	K	836	K	907	K	906	986
7	36	754	20	760	19	760	12	760	12	760	K	831	K	831	911
8	36	624	20	630	K	696	K	630	K	761	K	760	K	760	840
9	36	564	20	570	K	617	K	570	K	695	K	695	K	695	775
10	36	514	20	520	K	542	K	520	K	633	K	633	K	632	712
11	36	464	20	470	19	470	K	470	K	574	K	574	K	574	654
12	36	394	20	400	19	400	K	520	K	520	K	520	K	520	600
13	36	344	20	350	K	575	K	464	K	470	K	470	K	470	550
14	36	304	20	310	K	521	K	411	K	424	K	424	K	424	50
15	36	274	20	280	K	470	12	362	K	381	K	381	K	381	461
16	36	244	20	250	K	423	12	315	K	341	K	142	K	3442	422
17	36	224	20	230	K	380	12	271	K	306	K	306	K	306	386
18	36	204	20	210	K	338	12	230	K	273	K	273	K	273	353
19	36	184	20	190	K	300	12	190	K	242	K	243	K	243	323
20	36	169	20	280	K	264	12	175	K	214	K	214	K	215	295
21	K	876	K	213	K	229	12	160	K	188	K	189	K	189	269
22	K	801	K	145	K	197	12	145	K	164	K	165	K	166	246
23	K	728	20	130	K	166	12	130	K	144	K	144	K	144	224
24	K	658	20	120	K	136	12	120	K	143	K	125	K	126	206
25	K	592	20	110	19	110	12	110	K	109	12	110	K	111	191
26	K	530	20	100	19	100	12	100	K	97	12	100	12	100	180
27	K	469	20	30	19	90	12	90	12	90	12	90	12	90	170
28	K	412	20	80	19	80	12	80	12	80	12	80	12	80	160
29	K	356	20	70	19	70	12	70	12	70	12	70	12	70	150
30	K	306	20	65	19	65	12	65	12	65	12	65	12	65	145
31	K	261	20	60	19	60	12	60	12	60	12	60	12	60	140
32	K	218	20	55	19	55	12	55	12	55	12	55	12	55	135
33	K	176	20	50	19	50	12	50	12	50	12	50	12	50	130
34	K	140	20	40	19	40	12	40	12	40	12	40	12	40	150
35	K	111	20	35	19	35	12	35	12	35	12	35	12	35	115
36	K	84	20	30	19	30	12	30	12	30	12	30	12	30	110
37	K	55	20	25	19	25	12	25	12	25	12	25	12	25	105
38	K	33	20	15	19	15	12	15	12	15	12	15	12	15	95
39	K	15	20	7	19	7	12	7	12	7	12	7	12	7	87
40	K	0	20	0	19	0	12	0	12	0	12	0	12	0	80

(Howard, 1960)

Πίνακας 4: Αποτελέσματα Αντικατάστασης Αυτοκινήτου

Ο πίνακας περιλαμβάνει τις ακόλουθες στήλες:

1. State (Κατάσταση): Η ηλικία του αυτοκινήτου σε τριμηναία περίοδο
2. Επαναλήψεις 1 έως 7: Δείχνει τα αποτελέσματα για κάθε επανάληψη, τα οποία περιλαμβάνουν:
    - Κέρδος: Το κέρδος από την πολιτική αντικατάστασης.
    - Απόφαση: Η ενέργεια που πρέπει να ληφθεί, είτε να κρατηθεί το αυτοκίνητο ("K") είτε να αντικατασταθεί με ένα νέο ή παλαιότερο αυτοκίνητο (με τον αριθμό των τριμηναίων περιόδων).

- Αξία: Η αξία του αυτοκινήτου στην τρέχουσα κατάσταση.
- Προσαρμοσμένη Αξία: Η αξία προσαρμοσμένη με την προσθήκη της αξίας ανταλλαγής των $80.

Συγκεκριμένα φαίνεται ότι:

- Επανάληψη 1: Η αρχική πολιτική με κέρδος $250.00 και τις αντίστοιχες αποφάσεις και αξίες.
- Επανάληψη 2: Βελτιωμένη πολιτική με κέρδος $193.89, δείχνοντας αλλαγές στις αποφάσεις αντικατάστασης.
- Επανάληψη 3: Περαιτέρω βελτίωση με κέρδος $162.44.
- Επανάληψη 4: Το κέρδος αυξάνεται στα $157.07, με πιο προσεγμένες αποφάσεις.
- Επανάληψη 5: Κέρδος $151.05, διατηρώντας τις αποφάσεις πιο σταθερές.
- Επανάληψη 6: Ακόμα πιο σταθερή πολιτική με κέρδος $150.99.
- Επανάληψη 7: Τελική πολιτική με κέρδος $150.95, δείχνοντας ότι η πολιτική έχει σταθεροποιηθεί και οι αποφάσεις είναι βέλτιστες.

Από την ανάλυση του πίνακα μπορούμε να συμπεράνουμε ότι: Οι επαναλήψεις οδηγούν σε συνεχείς βελτιώσεις της πολιτικής αντικατάστασης. Το τελικό κέρδος και η πολιτική σταθεροποιούνται μετά την 7η επανάληψη. Οι αποφάσεις αντικατάστασης γίνονται πιο σαφείς και βέλτιστες με κάθε επανάληψη, ενισχύοντας την απόδοση και μειώνοντας τα έξοδα. Η πολιτική που προκύπτει από την τελευταία επανάληψη είναι η βέλτιστη και μπορεί να χρησιμοποιηθεί για την αποτελεσματική διαχείριση της συντήρησης και αντικατάστασης των αυτοκινήτων.

Η βέλτιστη πολιτική που προέκυψε από την επανάληψη 7 είναι να διατηρήσουμε στην κατοχή μας το αυτοκίνητο αν είναι μεταξύ 3 και 24 περιόδων. Αν το αυτοκίνητο είναι άλλης ηλικίας, να αντικατασταθεί με ένα αυτοκίνητο 12 περιόδων. Αυτή η πολιτική σχετίζεται με την λογική να διατηρούμε την ιδιοκτησία του αυτοκινήτου.

Η εξέλιξη του κόστους με τις επαναλήψεις παρατηρούμε ότι το κόστος ανά περίοδο μειώνεται εκθετικά με τις επαναλήψεις. Οι τελευταίες τρεις επαναλήψεις έχουν τόσο κοντινά κέρδη που οι αντίστοιχες πολιτικές μπορούν να θεωρηθούν ισοδύναμες. Σε ότι αφορά τη βέλτιστη αγορά η επιλογή ότι το καλύτερο αυτοκίνητο για αγορά είναι ηλικίας 3 ετών έγινε ήδη από την επανάληψη 4.

Τέλος συμπεραίνουμε ότι η βέλτιστη πολιτική όχι μόνο μειώνει το ετήσιο κόστος μεταφοράς αλλά και μεγιστοποιεί τη μελλοντική ανταμοιβή, αποδεικνύοντας την αποτελεσματικότητα της μεθόδου πολιτικής επανάληψης. Οι πολιτικές που προκύπτουν από τις τελευταίες επαναλήψεις είναι πρακτικά ισοδύναμες, υποδεικνύοντας τη σταθερότητα του συστήματος. Η μέθοδος πολιτικής επανάληψης μπορεί να εφαρμοστεί και σε άλλα βιομηχανικά προβλήματα αντικατάστασης με ανάλογη αποτελεσματικότητα. Αυτό το πλαίσιο παρέχει μια σαφή κατεύθυνση για τη λήψη αποφάσεων αντικατάστασης με βάση την ανάλυση κόστους-αξίας και τη μακροπρόθεσμη οικονομική στρατηγική.

ΚΕΦΑΛΑΙΟ 4

Μη επανδρωμένο Ελικόπτερο Αναζητά Ζώνη Προσγειώσεως σε Άγνωστο Περιβάλλον

4.1 Εισαγωγή

Οι (Sigaud & Baffet, 2010) στο βιβλίο τους παρουσιάζουν μια σημαντική πρόοδο στον τομέα της ρομποτικής και των Μη Επανδρωμένων Συστημάτων. Οι Sigaud & Baffet (2010) μελετούν ένα Ελικόπτερο έρευνας και διάσωσης σε ένα μερικώς γνωστό και αβέβαιο περιβάλλον, όπου αναζητά ζώνη προσγειώσεως, με την εφαρμογή των Μαρκοβιανών Διαδικασιών Αποφάσεων στο πρόβλημα της βελτιστοποίησης πολιτικής για την επίτευξη της αποστολής του. Σε αυτό το σενάριο, οι Μαρκοβιανές Διαδικασίες Αποφάσεων πλαισιώνουν τη λήψη αποφάσεων σε συνθήκες αβεβαιότητας, επιτρέποντας στο ελικόπτερο να κάνει επιλογές σε πραγματικό χρόνο σχετικά με τις ενέργειές του με βάση την τρέχουσα κατάσταση του περιβάλλοντος.

Το πρόβλημα μοντελοποιείται με τη χρήση των Μαρκοβιανών Διαδικασιών Αποφάσεων, όπου το ελικόπτερο λειτουργεί σε ένα περιβάλλον που αποτελείται από καταστάσεις, και σε κάθε κατάσταση, μπορεί να ληφθεί ένα σύνολο ενεργειών. Ο στόχος είναι να βρεθεί μια βέλτιστη πολιτική που υποδεικνύει ποιες ενέργειες πρέπει να ληφθούν σε κάθε κατάσταση για να μεγιστοποιηθεί η ανταμοιβή, η οποία θα μπορούσε να σχετίζεται με την εύρεση μιας ζώνης προσγείωσης ελαχιστοποιώντας παράλληλα τον κίνδυνο. Οι καταστάσεις αντιπροσωπεύουν την τρέχουσα κατάσταση του ελικοπτέρου και του περιβάλλοντος, συμπεριλαμβανομένης της θέσης του, του περιβάλλοντος του εδάφους, των γνωστών εμποδίων, των υπολειπόμενων καυσίμων, κ.λπ. Αυτές οι καταστάσεις αντιπροσωπεύονται συχνά σε μια δομημένη μορφή που συλλέγει σχετικές πληροφορίες για τη λήψη αποφάσεων. Οι ενέργειες που είναι διαθέσιμες στο ελικόπτερο περιλαμβάνουν συνήθως επιλογές κίνησης όπως εμπρός, πίσω, αριστερά και δεξιά, καθώς και ενέργειες υψηλότερου επιπέδου όπως η εξερεύνηση μιας συγκεκριμένης περιοχής ή η σάρωση για πιθανές ζώνες προσγείωσης.
(Sigaud & Baffet (2010))

Κάθε ενέργεια που αναλαμβάνεται από το ελικόπτερο μπορεί να οδηγήσει σε μετάβαση σε μια νέα κατάσταση με σχετικές πιθανότητες. Μια συνάρτηση ανταμοιβής ορίζεται για να ποσοτικοποιήσει το επιθυμητό να βρίσκεται σε μια συγκεκριμένη κατάσταση ή να εκτελεί μια συγκεκριμένη ενέργεια. Η συνάρτηση ανταμοιβής καταγράφει τους στόχους της αποστολής έρευνας και διάσωσης, όπως η εύρεση μιας ασφαλούς ζώνης προσγείωσης, η αποφυγή συγκρούσεων, η εξοικονόμηση ενέργειας κ.λπ. Το μοντέλο μετάβασης περιγράφει τα πιθανά αποτελέσματα των δράσεων σε διαφορετικές καταστάσεις. Δεδομένου ότι το περιβάλλον είναι μερικώς γνωστό και αβέβαιο, αυτές οι πιθανότητες μετάβασης μπορούν να εκτιμηθούν με βάση δεδομένα αισθητήρων, ιστορικές παρατηρήσεις ή μάθηση από την εμπειρία.

Οι διαδικτυακοί αλγόριθμοι βελτιστοποίησης, προσαρμοσμένοι στις Μαρκοβιανές Διαδικασίες Αποφάσεων, χρησιμοποιούνται για τον υπολογισμό της βέλτιστης πολιτικής σε πραγματικό χρόνο καθώς το ελικόπτερο πλοηγείται στο περιβάλλον. Αυτοί οι αλγόριθμοι ενημερώνουν επαναληπτικά την πολιτική με βάση νέες παρατηρήσεις και προσαρμόζουν ανάλογα τις ενέργειες του ελικοπτέρου. Η διαδικασία λήψης αποφάσεων που βασίζεται στις Μαρκοβιανές Διαδικασίες Αποφάσεων είναι στενά ενσωματωμένη με το σύστημα ελέγχου του ελικοπτέρου, το οποίο συλλέγει πληροφορίες για το περιβάλλον και τη μονάδα λήψης αποφάσεων, η οποία ερμηνεύει αυτές τις πληροφορίες για να δημιουργήσει ενέργειες. Αυτή η ενοποίηση επιτρέπει τον συντονισμό μεταξύ ανίχνευσης, λήψης αποφάσεων και εκτέλεσης ενεργειών. Ενώ το ελικόπτερο λειτουργεί αυτόνομα για το μεγαλύτερο μέρος της αποστολής, η ανθρώπινη παρέμβαση προορίζεται για κρίσιμες εργασίες, όπως η τελική επικύρωση πριν από την προσγείωση ή η διαχείριση ζητημάτων ασφαλείας κατά τη διάρκεια της πτήσης. Αυτό διασφαλίζει την ασφάλεια και παρέχει ένα επιπλέον επίπεδο επίβλεψης και ελέγχου. (Sigaud & Baffet (2010))

Αξιοποιώντας τις Μαρκοβιανές Διαδικασίες Αποφάσεων και τους διαδικτυακούς αλγόριθμους βελτιστοποίησης, το αυτόνομο ελικόπτερο έρευνας και διάσωσης μπορεί αποτελεσματικά να εξερευνήσει και να πλοηγηθεί σε πολύπλοκα και αβέβαια περιβάλλοντα, εκπληρώνοντας τελικά τους στόχους της αποστολής του, ελαχιστοποιώντας τους κινδύνους και την κατανάλωση πόρων. Αυτή η προσέγγιση αντιπροσωπεύει μια σημαντική πρόοδο στην ανάπτυξη μη επανδρωμένων συστημάτων για εφαρμογές στην πραγματικότητα όπως οι επιχειρήσεις έρευνας και διάσωσης. (Sigaud & Baffet (2010))

4.2 Ανάλυση Σεναρίου

Το σενάριο εξερεύνησης που περιγράφουν οι (Sigaud & Baffet, 2010) περιλαμβάνει δύο ελικόπτερα τύπου Yamaha RMAX εξοπλισμένα με αρχιτεκτονική ελέγχου ηλεκτρονικού εξοπλισμού και ενσωματωμένους αισθητήρες, που επιτρέπουν την αυτόνομη πτήση και εξερεύνηση σε μερικώς γνωστών περιβαλλόντων. Ο στόχος είναι να βρεθεί μια κατάλληλη ζώνη προσγείωσης για επιχειρήσεις έρευνας και διάσωσης.

Ακολούθως η ανάλυση του σεναρίου περιλαμβάνει την αρχική εγκατάσταση όπου τα ελικόπτερα παρέχονται με αρχικές πληροφορίες, συμπεριλαμβανομένης μιας θέσης GPS κοντά στην υποτιθέμενη τοποθεσία του ατόμου που πρόκειται να διασωθεί, απαγορευμένες ζώνες πτήσης και μια περιοχή αρχικής αναζήτησης που είναι εν μέρει γνωστή. Τα ελικόπτερα ξεκινούν πραγματοποιώντας μια αρχική εξερεύνηση σε υψόμετρο 50 μέτρων για τη δημιουργία ενός χάρτη της περιοχής με ληφθείσες εικόνες. Αυτή η χαρτογράφηση χωρίζει την περιοχή σε μικρότερες ζώνες που αντιστοιχούν σε μεγάλα εμπόδια ή πιθανές τοποθεσίες προσγείωσης. Αυτό το βήμα επιτρέπει μια πρόχειρη κατανόηση του περιβάλλοντος χωρίς να πλησιάσουν πολύ κοντά σε εμπόδια.

Μετά την αρχική χαρτογράφηση, τα ελικόπτερα πραγματοποιούν λεπτομερέστερη εξερεύνηση σε υψόμετρο 20 μέτρων αποφεύγοντας τα εμπόδια. Αυτή η φάση στοχεύει στον ακριβέστερο χαρακτηρισμό των εμποδίων και στην επιβεβαίωση πιθανών θέσεων προσγείωσης που εντοπίστηκαν κατά τη φάση της πρόχειρης εξερεύνησης. Ωστόσο, η στρατηγική εξερεύνησης πρέπει να λαμβάνει υπόψη την αυτονομία καυσίμου και τους περιορισμούς διάρκειας πτήσης.

Για την αποτελεσματική εξερεύνηση του περιβάλλοντος λαμβάνοντας υπόψη τους περιορισμούς καυσίμου και τους στόχους της αποστολής, τα ελικόπτερα χρησιμοποιούν έναν αλγόριθμο στοχαστικού σχεδιασμού ανά πάσα στιγμή. Αυτός ο αλγόριθμος δημιουργεί μια βελτιστοποιημένη πολιτική εξερεύνησης, παρέχοντας καθοδήγηση σχετικά με το ποιες επιμέρους ζώνες πρέπει να εξερευνηθούν, τις ενέργειες για να κινηθεί που πρέπει να εκτελεστούν και αποφάσεις σχετικά με την προσγείωση, την απογείωση και την επιστροφή στη βάση.

Η πολιτική εξερεύνησης ενημερώνεται συνεχώς με βάση πληροφορίες και περιορισμούς σε πραγματικό χρόνο. Αυτό επιτρέπει στα ελικόπτερα να προσαρμόζονται στις μεταβαλλόμενες συνθήκες και να λαμβάνουν αποφάσεις δυναμικά καθώς εξερευνούν το περιβάλλον.

Συνολικά, αυτό το σενάριο απεικονίζει τη χρήση προηγμένων αυτόνομων συστημάτων και αλγορίθμων σχεδιασμού για να επιτρέψουν στα Μη επανδρωμένα ελικόπτερα να εξερευνήσουν μερικώς γνωστά περιβάλλοντα, να εντοπίσουν πιθανές τοποθεσίες προσγείωσης και να εκτελέσουν αποστολές έρευνας και διάσωσης αποτελεσματικά και με ασφάλεια (Sigaud & Baffet, 2010).

4.3 Πρόβλημα Σχεδιασμού

Οι (Sigaud & Baffet, 2010) στο πρόβλημα σχεδιασμού που προκύπτει μετά την αρχική χαρτογράφηση της περιοχής περιλαμβάνει τη δημιουργία ενός σχεδίου εξερεύνησης για τον περαιτέρω χαρακτηρισμό των προσδιορισμένων επιμέρους ζωνών και τον προσδιορισμό της καταλληλόλητας τους για προσγείωση. Αυτό το πρόβλημα σχεδιασμού διατυπώνεται χρησιμοποιώντας τον στοχαστικό σχεδιασμό, μια γλώσσα που έχει σχεδιαστεί για την περιγραφή πιθανών καταστάσεων και προβλημάτων στον τομέα της τεχνητής νοημοσύνης, ειδικότερα στον τομέα του σχεδιασμού συστημάτων που λειτουργούν σε περιβάλλοντα με αβεβαιότητα. Ο σχεδιασμός γίνεται με χρήση συγκεκριμένης γλώσσας μετά από κάποια αρχική χαρτογράφηση κάποιας περιοχής. Ο χάρτης θα δώσει πληροφορίες για τις επιμέρους ζώνες που έχουν συντεταγμένες, διάσταση, αριθμό σημείων διαδρομής, ακόμη και την πιθανότητα καταλληλόλητας τους για προσγείωση. Χρησιμοποιώντας τη γλώσσα προγραμματισμού, το περιβάλλον περιγράφει το πρόβλημα που πρέπει να λυθεί με καταστάσεις, προϋποθέσεις και αποτελέσματα των ενεργειών που λαμβάνουν μία λογική συνθήκη πχ.« Μετακίνηση Πίσω». Επίσης θα δώσει μια περιοχή εντός της χαρτογραφημένης επιμέρους ζώνης με ένα συγκεκριμένο όνομα και όλες τις δισδιάστατες συντεταγμένες της, τις διαστάσεις, τον αριθμό των σημείων διαδρομής και την πιθανότητα να είναι κατάλληλη για προσγείωση.

Οι τρεις κύριες ενέργειες που εξηγούνται σε αυτόν τον αλγόριθμο είναι «Πήγαινε στη ζώνη προσγείωσης», «προσγείωση» και «απογείωση». Πιθανότατα, αυτές οι τρεις ενέργειες αναφέρονται στη μετακίνηση σε κάποια επιμέρους ζώνη, στην προσγείωση σε κάποια επιμέρους ζώνη και στην απογείωση από κάποια επιμέρους ζώνη, αντίστοιχα. Εκμεταλλεύεται τη Boolean Αληθές-Ψευδές, αντιπροσωπεύοντας χαρακτηριστικά του τομέα σχεδιασμού, για παράδειγμα, χαρακτηριστικά γνωρίσματα επιμέρους ζωνών. Συνοπτικά, αυτή η γλώσσα σχεδιασμού χρησιμοποιείται στον σχεδιασμό του προβλήματος εξερεύνησης αφού έχει γίνει πρόχειρη χαρτογράφηση μιας περιοχής, η οποία δίνει συστηματικό και λογικό σχεδιασμό ενεργειών και λήψη αποφάσεων με βάση τα χαρακτηριστικά από τις επιμέρους ζώνες. (Sigaud & Baffet, 2010)

4.4 Αβεβαιότητα

Όπως αναφέρουν οι Fabiani & Teichteil-Konigsbuch (2007), η βελτιστοποίηση της πολιτικής στο περιγραφόμενο σενάριο περιλαμβάνει την εξέταση τριών πηγών αβεβαιότητας:

Πιθανότητα Καταλληλότητας επιμέρους ζώνης:

$$(\text{Pa}) = \frac{\text{αριθμός εικονοστοιχείων φωτός}}{\text{αριθμός σκούρων εικονοστοιχείων}}$$

Όπου ο αριθμός των εικονοστοιχείων μετράται μετά την επεξεργασία της εικόνας [τα εικονοστοιχεία μετά την επεξεργασία εικόνας]

Αυτή η πιθανότητα αντιπροσωπεύει την πιθανότητα ότι μια επιμέρους ζώνη που προσδιορίζεται ως κατάλληλη για προσγείωση κατά την αρχική πρόχειρη χαρτογράφηση θα επιβεβαιωθεί πράγματι ως κατάλληλη μετά από περαιτέρω χαρακτηρισμό.

Υπολογίζεται με βάση την αναλογία φωτεινών εικονοστοιχείων (που υποδεικνύει την καταλληλόλητα) προς τα σκοτεινά εικονοστοιχείων (που υποδηλώνει ακαταλληλότητα) μετά την επεξεργασία της εικόνας.

Πιθανότητα επιτυχούς διάσωσης (Ps):

$$P_s = \frac{40}{40 + d_z}$$

Αυτή η πιθανότητα αντιπροσωπεύει την πιθανότητα επιτυχούς διάσωσης ενός ανθρώπου εάν το ελικόπτερο προσγειωθεί σε μια ορισμένη απόσταση (d_z) από τη θέση του ανθρώπου μέσα σε μια επιμέρους ζώνη.

Υπολογίζεται με βάση έναν τύπο όπου το P_s αυξάνεται καθώς η απόσταση (d_z) μειώνεται, υποδεικνύοντας μεγαλύτερη πιθανότητα επιτυχίας όταν το ελικόπτερο προσγειώνεται πιο κοντά στον άνθρωπο.

Πυκνότητα πιθανότητας ενέργειας

$$f_r = \frac{1}{\sigma_\alpha \sqrt{2\pi}} e^{-\frac{(r-\mu_\alpha)^2}{2\sigma^2}}$$

Αυτή η συνάρτηση πυκνότητας πιθανότητας αντιπροσωπεύει την πιθανότητα μιας ενέργειας που διαρκεί μια ορισμένη διάρκεια (t) σε δευτερόλεπτα (Fabiani & Teichteil-Konigsbuch (2007))

Μοντελοποιείται χρησιμοποιώντας την κανονική κατανομή με μ_a και σ_a αντιπροσωπεύουν τη μέση και τυπική απόκλιση της διάρκειας ενέργειας για μια συγκεκριμένη ενέργεια (a). Αυτές οι πηγές αβεβαιότητας είναι ζωτικής σημασίας για τη βελτιστοποίηση της πολιτικής, καθώς παρέχουν πιθανά μέτρα για τη λήψη αποφάσεων. Με την ενσωμάτωση αυτών των πιθανοτήτων στη διαδικασία σχεδιασμού, η πολιτική μπορεί να βελτιστοποιηθεί για να μεγιστοποιηθεί η πιθανότητα επιτυχόντων αποτελεσμάτων, όπως η επιβεβαίωση κατάλληλων ζωνών προσγείωσης, η αποτελεσματική διάσωση ανθρώπων και η ακριβής εκτίμηση της διάρκειας ενέργειας.

4.5 Το κριτήριο βελτιστοποίησης

Ως κριτήριο βελτιστοποίησης εννοούμε τη μαθηματική προσδοκώμενη τιμή του συνολικού αθροίσματος όλων των ανταμοιβών ή ποινών που λαμβάνονται μετά την εκτέλεση των ενεργειών. Ορίζεται ως ανταμοιβή +1000 που σχετίζεται με τη διάσωση του ανθρώπου (όταν η μεταβλητή "human-rescued" είναι TRUE και το ελικόπτερο έχει προσγειωθεί με ασφάλεια). Ενώ ορίζεται ως «ποινή» -1000 αντιστοιχεί στην επιστροφή του ελικοπτέρου στη βάση του χωρίς να διασώσει τον άνθρωπο.

Προκειμένου να ληφθεί η απόφαση για την καταλληλότερη ζώνη προσγειώσεως, πρέπει να ληφθούν υπόψη οι πιθανότητες για κάθε επιμέρους ζώνη να είναι εν τέλει κατάλληλη για προσγείωση, οι πιθανότητες διάσωσης του ανθρώπου με τη προσγείωση σε αυτήν την επιμέρους ζώνη και τις πιθανότητες επιστροφής στην αρχική βάση με ασφάλεια από αυτήν την επιμέρους ζώνη. Η λήψη αποφάσεων λαμβάνει υπόψη όλες τις πιθανότητες για να καθορίσει την καλύτερη στρατηγική δράσης, μεγιστοποιώντας έτσι την αναμενόμενη συνολική ανταμοιβή και ελαχιστοποιώντας τις πιθανές ποινές.

Η πολιτική σύμφωνα με τους (Sigaud & Baffet, 2010) υπολογίζεται για κάθε επιμέρους ζώνη. Η αναμενόμενη ανταμοιβή, η οποία είναι το άθροισμα των ανταμοιβών και των ποινών που λαμβάνονται μετά την εκτέλεση ενεργειών. P_a είναι η πιθανότητα η επιμέρους ζώνη να είναι κατάλληλη για προσγείωση. Το P_s είναι η πιθανότητα επιτυχούς διάσωσης του ανθρώπου εάν το ελικόπτερο προσγειωθεί στην επιμέρους ζώνη. Pr_safe(z) είναι η πιθανότητα ασφαλούς επιστροφής στο σπίτι από την επιμέρους ζώνη.1000 είναι η ανταμοιβή για τη διάσωση του ανθρώπου και (−1000) είναι η ποινή για την επιστροφή στη βάση χωρίς τη διάσωση του ανθρώπου. Η ταξινόμηση των επιμέρους ζωνών γίνεται με βάση την αναμενόμενη χρησιμότητά τους και δίνει προτεραιότητα στις επιμέρους ζώνες με την υψηλότερη αναμενόμενη χρησιμότητα, υποδεικνύοντας υψηλότερες πιθανότητες επιτυχόντων αποτελεσμάτων.

Βελτιστοποιώντας τη πολιτική με αυτόν τον τρόπο και λαμβάνοντας υπόψη τις πιθανότητες που σχετίζονται με την καταλληλότητα προσγείωσης, την επιτυχή διάσωση και την ασφαλή επιστροφή, η διαδικασία εξερεύνησης μεγιστοποιεί την αναμενόμενη χρησιμότητα, αυξάνοντας έτσι την πιθανότητα επίτευξης των επιθυμητών αποτελεσμάτων, ελαχιστοποιώντας τον κίνδυνο αποτυχίας.

4.6 Η ενσωματωμένη Αρχιτεκτονική Ελέγχου και Αποφάσεων

Στην παρούσα ενότητα αναλύεται η αρχιτεκτονική ελέγχου και η διαδικασία λήψης απόφασης (Sigaud & Baffet, 2010) περιλαμβάνει δύο κύρια επίπεδα: το επίπεδο αποφάσεων (deliberative layer) και το επίπεδο ανατροφοδότησης (reactive layer).

Το επίπεδο ελέγχου: είναι υπεύθυνο για τη λήψη αποφάσεων και τον προγραμματισμό υψηλού επιπέδου. Αποτελείται από τρεις κύριες λειτουργίες:

Την όραση η οποία επεξεργάζεται τις αποκτηθείσες εικόνες, προσδιορίζει τις επιμέρους ζώνες και τις χαρακτηρίζει. Αυτή η λειτουργία παρέχει ουσιαστικά στοιχεία για τον προγραμματισμό. Τον σχεδιασμό που αφορά την εύρεση της βέλτιστης στρατηγικής με τη χρήση των πλήρως Παρατηρήσιμων Μαρκοβιανών Διαδικασιών αποφάσεων με βάση την έξοδο της συνάρτησης όρασης. Στη συνέχεια σχεδιάζεται η σειρά των ενεργειών για την επίτευξη των στόχων της αποστολής μεγιστοποιώντας τις ανταμοιβές και ελαχιστοποιώντας τις ποινές. Ο επόπτης εκτελεί αυτόματα την επίβλεψη φάσεων αποστολής, συντονίζει τις λειτουργίες όρασης και προγραμματισμού και στέλνει εντολές ελέγχου στο επίπεδο ανατροφοδότησης για εκτέλεση (Sigaud & Baffet, 2010).

Το επίπεδο ανατροφοδότησης είναι υπεύθυνο για λειτουργίες ελέγχου πτήσης χαμηλού επιπέδου και λειτουργεί υπό περιορισμούς σε πραγματικό χρόνο. Εξασφαλίζει την ασφάλεια και τη σταθερότητα της πτήσης εκτελώντας επικυρωμένες λειτουργίες ελέγχου πτήσης ανεξάρτητα από το επίπεδο ελέγχου.

Ο διαχωρισμός των επιπέδων ελέγχου και ανατροφοδότησης επιτρέπει στο σύστημα να εξισορροπεί αποτελεσματικά τη μνήμη και τους υπολογιστικούς πόρους. Το επίπεδο Ανατροφοδότησης δίνει προτεραιότητα στον έλεγχο πτήσης σε πραγματικό χρόνο, ενώ το επίπεδο ελέγχου εστιάζει στη λήψη αποφάσεων υψηλότερου επιπέδου χωρίς να επηρεάζει την ασφάλεια της πτήσης. Ενώ το επίπεδο ελέγχου μπορεί να καταναλώνει περισσότερη μνήμη και υπολογιστικό χρόνο, πρέπει να παράγει αποφάσεις εντός εύλογου χρονικού πλαισίου για την αποφυγή επικίνδυνων καταστάσεων.

Η αρχιτεκτονική του συστήματος στοχεύει να παρέχει μια συμπεριφορά "ανά πάσα στιγμή", επιτρέποντας το επίπεδο ελέγχου να δημιουργεί ασφαλείς ενέργειες αμέσως, ακόμα κι αν δεν έχει βελτιστοποιήσει πλήρως το σχέδιο αποστολής. Συνολικά, αυτή η αρχιτεκτονική διευκολύνει την αποτελεσματική λήψη αποφάσεων και τον συντονισμό μεταξύ της αντίληψης, του σχεδιασμού και της εκτέλεσης ενέργειας, επιτρέποντας στα αυτόνομα συστήματα στροφείου να λειτουργούν με ασφάλεια και αποτελεσματικότητα σε πολύπλοκα και αβέβαια περιβάλλοντα. Στην ενσωματωμένη αρχιτεκτονική που περιγράφεται, η αποστολή εποπτεύεται από ένα κεντρικό στοιχείο γνωστό ως επόπτης. Αυτός ο επόπτης ενεργεί ως συντονιστής και ελεγκτής, επιβλέποντας τη λειτουργία ολόκληρου του συστήματος. Ο επόπτης αλληλεπιδρά με το σύστημα απεικόνισης και του προγράμματος και χρησιμεύει ως το κεντρικό στοιχείο της αρχιτεκτονικής, υπεύθυνος για την επίβλεψη της εκτέλεσης της αποστολής. Συντονίζει τις δραστηριότητες άλλων εξαρτημάτων και διασφαλίζει την ομαλή λειτουργία του συστήματος. Ο επόπτης επικοινωνεί τόσο με το σύστημα απεικόνισης όσο και με τον σχεδιαστή για να συγκεντρώσει τις απαραίτητες πληροφορίες και να λάβει αποφάσεις σχετικά με την αποστολή. Το σύστημα απεικόνισης είναι υπεύθυνο για τη λήψη εικόνων του περιβάλλοντος και την επεξεργασία τους για την εξαγωγή σχετικών πληροφοριών. Παρέχει στον επόπτη δεδομένα που σχετίζονται με την τρέχουσα κατάσταση του περιβάλλοντος, όπως οι θέσεις των εμποδίων, οι πιθανές ζώνες προσγείωσης και άλλα σχετικά χαρακτηριστικά. Ο σχεδιαστής είναι υπεύθυνος για τη δημιουργία και τη βελτιστοποίηση του σχεδίου αποστολής με βάση τις πληροφορίες που παρέχονται από το σύστημα απεικόνισης. Λαμβάνει πληροφορίες από τον επόπτη σχετικά με τους στόχους και τους περιορισμούς της αποστολής και δημιουργεί ένα σχέδιο που περιγράφει τη σειρά των ενεργειών που πρέπει να γίνουν για την επίτευξη αυτών των στόχων. Ο σχεδιαστής μπορεί να χρησιμοποιήσει διάφορους αλγόριθμους βελτιστοποίησης για να διασφαλίσει ότι το σχέδιο αποστολής είναι αποδοτικό και αποτελεσματικό. (Fabiani & Teichteil-Konigsbuch (2007)).

Ο επόπτης ενσωματώνει πληροφορίες τόσο από το σύστημα απεικόνισης όσο και από τον σχεδιαστή για να λάβει αποφάσεις σχετικά με την πρόοδο της αποστολής, διασφαλίζοντας ότι το σύστημα λειτουργεί αποτελεσματικά και επιτυγχάνει τους στόχους του.

Συνολικά, ο επόπτης διαδραματίζει κρίσιμο ρόλο στο συντονισμό των δραστηριοτήτων του συστήματος απεικόνισης και του σχεδιαστή, επιτρέποντας στο αυτόνομο σύστημα να εκτελεί την αποστολή του αποτελεσματικά σε δυναμικά και αβέβαια περιβάλλοντα.

Στην ανάλυση τους οι Fabiani & Teichteil-Konigsbuch (2007) τονίζουν ότι η ενσωματωμένη αρχιτεκτονική ελέγχου και αποφάσεων είναι ένα πλαίσιο που έχει σχεδιαστεί για να διευκολύνει την αυτόνομη λήψη αποφάσεων και τον έλεγχο σε μη επανδρωμένα συστήματα, όπως τα αυτόνομα εναέρια οχήματα. Αυτή η αρχιτεκτονική συνήθως αποτελείται από πολλά επίπεδα και στοιχεία που συνεργάζονται για να επιτρέψουν στο σύστημα να αντιλαμβάνεται, να επεξεργάζεται, να σχεδιάζει και να εκτελεί εργασίες αυτόνομα. Το επίπεδο ελέγχου είναι υπεύθυνο για τη συλλογή δεδομένων σχετικά με το περιβάλλον μέσω διαφόρων αισθητήρων, όπως κάμερες, light detection & Ranging, ραντάρ ή άλλες μεθόδους ανίχνευσης. Επεξεργάζεται ακατέργαστα δεδομένα αισθητήρων για να εξάγει σχετικές πληροφορίες για το περιβάλλον, συμπεριλαμβανομένου του εδάφους, των εμποδίων, των στόχων και άλλων σχετικών χαρακτηριστικών. Το επίπεδο ελέγχου επεξεργάζεται τις πληροφορίες που παρέχονται από το επίπεδο αντίληψης και δημιουργεί αποφάσεις και σχέδια υψηλού επιπέδου με βάση τους στόχους και τους περιορισμούς της αποστολής.

Λαμβάνει υπόψη παράγοντες όπως τους στόχους της αποστολής, τις περιβαλλοντικές συνθήκες, τους περιορισμούς πόρων, τις απαιτήσεις ασφάλειας και τις προτεραιότητες της αποστολής κατά τη λήψη αποφάσεων.

Διασφαλίζει ότι οι ενέργειες του συστήματος εκτελούνται με ασφάλεια και αποτελεσματικότητα, λαμβάνοντας υπόψη την ανάδραση των αισθητήρων σε πραγματικό χρόνο και τη δυναμική του έλεγχο κίνησης, παρακολούθηση διαδρομής και αποφυγή εμποδίων. Οι μονάδες επικοινωνίας και ολοκλήρωσης διευκολύνουν την ανταλλαγή πληροφοριών και εντολών μεταξύ διαφορετικών στοιχείων της αρχιτεκτονικής και εξασφαλίζουν τη συνεργασία μεταξύ των διαδικασιών αντίληψης, λήψης αποφάσεων και ελέγχου.

Τα πρωτόκολλα και οι διεπαφές επικοινωνίας έχουν δημιουργηθεί για να επιτρέπουν τη διαλειτουργικότητα μεταξύ στοιχείων υλικού και λογισμικού. Τα στοιχεία εποπτείας και παρακολούθησης επιβλέπουν τη λειτουργία του συστήματος, διασφαλίζοντας ότι λειτουργεί εντός ασφαλών και νομικών ορίων.

περιβάλλοντος. Το επίπεδο ελέγχου μπορεί να περιλαμβάνει μονάδες για σχεδιασμό τροχιάς, να περιλαμβάνει συστήματα παρακολούθησης της υγείας, ελέγχους ασφαλείας, μηχανισμούς ανίχνευσης και ανάκτησης σφαλμάτων και παρακολούθηση της απόδοσης σε πραγματικό χρόνο. Οι μηχανισμοί ανάδρασης επιτρέπουν στο σύστημα να διδαχθεί από την εμπειρία και να προσαρμόσει

τη συμπεριφορά του με την πάροδο του χρόνου. Αυτοί οι μηχανισμοί μπορεί να περιλαμβάνουν διαδικτυακούς αλγόριθμους μάθησης, προσαρμοστικές στρατηγικές ελέγχου ή τεχνικές ενίσχυσης μάθησης (Fabiani & Teichteil-Konigsbuch (2007).

Οι Fabiani και Teichteil-Konigsbuch (2007) προσφέρουν μια σημαντική συμβολή στη βελτιστοποίηση στρατηγικών για την εξερεύνηση και διάσωση, αξιοποιώντας τις δυνατότητες των Μαρκοβιανών Διαδικασιών Αποφάσεων για τη λήψη αποφάσεων σε σύνθετα και δυναμικά περιβάλλοντα.

4.7 Λειτουργία του Επόπτη

Οι Fabiani & Teichteil-Konigsbuch (2007) με ένα διάγραμμα επεξηγούν τα στοιχεία της αρχιτεκτονικής απόφασης και ελέγχου, δείχνοντας τη σχέση μεταξύ της λειτουργίας προγραμματισμού και του επόπτη. Τα κύρια στοιχεία που εμφανίζονται σε σχέση μεταξύ τους σύμφωνα με τους (Sigaud & Baffet, 2010) είναι τα παρακάτω:

Έξοδος λειτουργίας όρασης: Στη συνάρτηση όρασης, η επιφάνεια που λαμβάνεται υφίσταται επεξεργασία και η χαρτογράφηση με βάση την εικόνα δημιουργείται για κάθε εικόνα.

Ο επόπτης επεξεργάζεται αυτόματα το πρόβλημα του στοχαστικού σχεδιασμού, δεδομένου του αποτελέσματος από τη λειτουργία όρασης. Ενεργοποιεί τον διακομιστή προγραμματισμού και στέλνει το πρόβλημα προγραμματισμού για επίλυση. Ο διακομιστής σχεδιασμού είναι μια υπηρεσία πολλαπλών διεργασιών που λαμβάνει το πρόβλημα σχεδιασμού από τον επόπτη. Εκκινεί δύο παράλληλες εργασίες. Την Βελτιστοποίηση της πολιτικής ενέργειας όπου η πολιτική δράσης σε σχέση με το πρόβλημα του στοχαστικού σχεδιασμού βελτιστοποιείται χρησιμοποιώντας αυτήν την δευτερεύουσα εργασία. Την Επικοινωνία με τον επόπτη παρέχοντας παράλληλα ενημέρωση σχετικά με την πρόοδο ή οποιαδήποτε άλλη απαιτούμενη πληροφορία. (Sigaud & Baffet, 2010)

4.8 Βελτιστοποίηση της πολιτικής

Η διαδικασία βελτιστοποίησης πολιτικής που περιγράφεται περιλαμβάνει σταδιακή δημιουργία και βελτίωση της πολιτικής σε διαδοχικές επαναλήψεις. Η πολιτική δημιουργείται σταδιακά μέσω διαδοχικών επαναλήψεων βελτιστοποίησης. Κάθε επανάληψη στοχεύει να βελτιώσει τοπικά την τρέχουσα πολιτική εστιάζοντας σε έναν σταδιακά διευρυμένο επιμέρους χώρο της κατάστασης. Κατά τη διάρκεια κάθε επανάληψης, ο στόχος είναι να βελτιωθεί η τρέχουσα πολιτική μέσα σε έναν επιμέρους χώρο προσβάσιμων καταστάσεων. Αυτός ο επιμέρους χώρος περιλαμβάνει την τρέχουσα κατάσταση και τις καταστάσεις στόχου, οι οποίες είναι καταστάσεις που παρέχουν θετική ανταμοιβή όταν επιτευχθεί.

Η τρέχουσα πολιτική και ο αντίστοιχος υποχώρος των προσβάσιμων καταστάσεων αποθηκεύονται σε μια ασφαλή θέση μνήμης. Μια ισχύουσα πολιτική είναι διαθέσιμη από το τέλος της πρώτης επανάληψης βελτιστοποίησης. Ο επόπτης μπορεί να εκτελέσει την πρώτη ενέργεια πριν να ολοκληρωθεί όλη η διαδικασία βελτιστοποίησης.

Η συνάρτηση σχεδιασμού σε οποιαδήποτε στιγμή μπορεί να βελτιώνει συνεχώς την πολιτική με την πάροδο του χρόνου. Η διαδικασία βελτιστοποίησης έχει σχεδιαστεί ώστε να είναι αρκετά αποτελεσματική ώστε να παρέχει μια εφαρμοστέα πολιτική εντός εύλογου υπολογιστικού χρόνου. Αυτό επιτρέπει στο σύστημα να λαμβάνει αποφάσεις και να αναλαμβάνει ενέργειες σε πραγματικό χρόνο, ακόμη και όταν η διαδικασία βελτιστοποίησης συνεχίζεται στο παρασκήνιο. Συνολικά, αυτή η προσέγγιση επιτρέπει στο σύστημα να βελτιώνει σταδιακά και να βελτιώνει την πολιτική λήψης αποφάσεων, διασφαλίζοντας παράλληλα ότι παραμένει λειτουργικό και ικανό να ανταποκρίνεται σε πραγματικές καταστάσεις έγκαιρα. (Fabiani & Teichteil-Konigsbuch (2007))

4.9 Στοχαστικός Δυναμικός Προγραμματισμός

Στο στοχαστικό δυναμικό προγραμματισμό, ο αλγόριθμος Stochastic Focused Dynamic Programming, χρησιμοποιείται για την επίτευξη βελτιστοποιήσεων των Μαρκοβιανών Διαδικασιών Αποφάσεων εντός των περιορισμών της λήψης αποφάσεων σε πραγματικό χρόνο. Στη λειτουργία αυτού του αλγοριθμικού πλαισίου οι (Sigaud & Baffet, 2010) παρουσιάζουν την εφαρμογή των ελικοπτέρων ReSSAC, τις συνθήκες ασφαλούς λειτουργίας οι οποίες απαιτούν τη λειτουργία σχεδιασμού που παράγει μια απόφαση εντός χρονικού πλαισίου συγκρίσιμου με τον χρόνο εκτέλεσης μιας ενέργειας UAV, που είναι κατά μέσο όρο περίπου 50 δευτερόλεπτα.

Ο στόχος περιλαμβάνει συνήθως την επίτευξη μιας συγκεκριμένης τοποθεσίας (π.χ. βάση), την εξασφάλιση ανθρώπινης διάσωσης ή τη διατήρηση μιας ελάχιστης αυτονομίας πτήσης.

Ο αλγόριθμος εναλλάσσεται μεταξύ δύο σταδίων σε κάθε επανάληψη βελτιστοποίησης. Το πρώτο στάδιο υπολογίζει το σύνολο των καταστάσεων που είναι προσβάσιμες από την τρέχουσα κατάσταση εφαρμόζοντας επαναληπτικά την τρέχουσα πολιτική μέχρι την επίτευξη των καθορισμένων καταστάσεων στόχου. Το δεύτερο στάδιο διεξάγει μια τοπική βελτιστοποίηση (ενημέρωση Bellman) της πολιτικής εντός του ληφθέντος συνόλου προσβάσιμων καταστάσεων.

Μεταξύ των επαναλήψεων βελτιστοποίησης, το μέγεθος του επιμέρους χώρου των προσβάσιμων καταστάσεων επεκτείνεται σταδιακά, επιτρέποντας πιο ολοκληρωμένη εξερεύνηση του χώρου καταστάσεων. Μόλις ληφθεί μια αρχική πολιτική, η τρέχουσα βέλτιστη πολιτική μπορεί να εκτελεστεί ανά πάσα στιγμή κατόπιν αιτήματος του επόπτη. Αυτή η δυνατότητα διασφαλίζει ότι μια ασφαλής ενέργεια είναι άμεσα διαθέσιμη για εκτέλεση, ακόμη και πριν ολοκληρωθεί η διαδικασία βελτιστοποίησης. Επιλέγεται αλγόριθμος στοχαστικού πεπερασμένου δυναμικού προγραμματισμού για την καταλληλότητα του σε στάδια επέκτασης και τοπικής βελτιστοποίησης, διευκολύνοντας την αποτελεσματική λήψη αποφάσεων εντός των καθορισμένων χρονικών περιορισμών.

Συνολικά, ο αλγόριθμος στοχαστικού πεπερασμένου δυναμικού προγραμματισμού παρέχει ένα αποτελεσματικό πλαίσιο για την επίτευξη βελτιστοποιήσεων των Μαρκοβιανών Διαδικασιών Αποφάσεων σε πραγματικό χρόνο, επιτρέποντας στο αυτόνομο σύστημα να λαμβάνει έγκαιρες και τεκμηριωμένες αποφάσεις σε δυναμικά και αβέβαια περιβάλλοντα.

4.10 Αρχική ασφαλής πολιτική

Στο κεφάλαιο αυτό ακολουθεί ο αλγόριθμος που αναφέρουν οι Fabiani & Teichteil-Konigsbuch (2007) για το πώς να καταλήξουμε γρήγορα σε μια αρχική ασφαλή πολιτική, αυτή που εγγυάται την ύπαρξη τουλάχιστον μιας τροχιάς από την αρχική κατάσταση σε μια κατάσταση στόχου. Το αλγοριθμικό σχήμα δίνεται παρακάτω:

$$\pi_0(s) = \begin{cases} \{\textit{ενέργειες οι οποίες ορίζονται από τις καταστάσεις στόχους}\} & \text{αν s } \in \text{ στόχος,} \\ \varnothing & \text{αλλιώς} \end{cases}$$

$$\pi_{t+1}(s) = \{a : \exists s', T(s'|a,s) > 0 \textit{ και } \pi_t(s')\}$$

Εάν η κατάσταση s είναι μια κατάσταση στόχου, τότε αρχικοποιείται η $\pi_0(s)$ με την πολιτική που περιέχει ενέργειες από τις καταστάσεις στόχου διαφορετικά, αρχικοποιεί το $\pi_0(s)$ με ένα κενό σύνολο. (Sigaud & Baffet, 2010)

Επανάληψη τα διαδοχικά χρονικά βήματα t για τη βελτίωση της πολιτικής.

Σε κάθε επανάληψη, ενημερώνεται η πολιτική $\pi_{t+1}(s)$ με κριτήριο ότι για κάθε κατάσταση s είναι μια κατάσταση στόχου και το $\pi_{t+1}(s)$ παραμένει αμετάβλητο.

Διαφορετικά, για κάθε κατάσταση s, θεωρείται ότι όλες οι ενέργειες a για τις οποίες υπάρχει κατάσταση s', με μη μηδενική πιθανότητα μετάβασης T(s'|s, a), και $\pi_t(s')$ δεν είναι κενή. Οι ενέργειες προστίθενται στο $\pi_{t+1}(s)$.

Λήγει όταν υπάρξει ένα βήμα t έτσι ώστε το π_t (αρχική κατάσταση) να μην είναι μηδέν, που σημαίνει ότι υπάρχει τουλάχιστον μία διαθέσιμη ενέργεια στην αρχική κατάσταση.

Επιπλέον, διασφαλίζει ότι οι επιλεγμένες ενέργειες αποτελούν ένα είδος της συντομότερης διαδρομής μεταξύ της τρέχουσας κατάστασης και των καταστάσεων στόχου. Αυτό γίνεται προτιμώντας ενέργειες που αντιστοιχούν σε τροχιές με λιγότερα βήματα από την τρέχουσα κατάσταση.

Η πρώτη επανάληψη υπολογίζει την πολιτική εκτελώντας κάποια λογικά τεστ, που

επιτρέπουν την επιλογή της πρώτης εφαρμοστέας ενέργειας, ανεξάρτητα από την τρέχουσα κατάσταση. Η διακοπή της διαδικασίας δυναμικού προγραμματισμού εφόσον βρεθεί μια σχετική ενέργεια από την κατάσταση έναρξης θα σήμαινε ότι υπάρχει μια πολιτική για την επίτευξη τουλάχιστον μιας τροχιάς προς μια κατάσταση στόχου. Γενικά, ο προτεινόμενος αλγόριθμος είναι γρήγορος στη δημιουργία μιας ασφαλούς αρχικής πολιτικής χωρίς πλήρη βελτιστοποίηση. Βασικά, ο αλγόριθμος κατασκευάζει μια πολιτική λαμβάνοντας υπόψη τη σωστή προτεραιότητα παραχώρησης στις ενέργειες που αντιστοιχούν σε μικρότερες διαδρομές και έτσι διασφαλίζει την ύπαρξη τουλάχιστον ενός μονοπατιού από την αρχική κατάσταση σε κάποια κατάσταση στόχου (Sigaud & Baffet, 2010).

4.11 Τα Αποτελέσματα Των Δοκιμών Πτήσης

Οι Fabiani & Teichteil-Konigsbuch (2007) πραγματοποίησαν δοκιμές πτήσης σε πραγματικό χρόνο και δημόσιες επιδείξεις της εφαρμοσμένης αρχιτεκτονικής απόφασης και των αλγορίθμων σχεδιασμού όπως παρουσίασαν οι (Sigaud & Baffet, 2010). Οι δοκιμές περιλάμβαναν σενάρια που αφορούσαν τεχνητή αντίληψη και αυτόνομο διαδικτυακό επανασχεδιασμό και διεξήχθησαν περαιτέρω δοκιμές απόδοσης. Η απόδοση αξιολογήθηκε με βάση διάφορα κριτήρια όπως ο συνολικός χρόνος βελτιστοποίησης, συμπεριλαμβανομένων των διαδικασιών αρχικού σχεδιασμού και επανασχεδιασμού. Μέγιστος χρόνος ανταπόκρισης, ο οποίος είναι ο χρόνος που απαιτείται για την αποστολή μιας ενέργειας στον επόπτη μετά τη λήψη της τρέχουσας κατάστασης. Άμεση σύνδεση ζώνης προσγείωσης με την εφαρμογή της τρέχουσας πολιτικής από τη βάση του UAV.

Το πρόβλημα σχεδιασμού περιελάμβανε 24 μεταβλητές κατάστασης και ένα μεγάλο χώρο κατάστασης άνω των 13 δισεκατομμυρίων καταστάσεων.

Οι μεταβλητές κατάστασης περιλάμβαναν παράγοντες όπως η κατάσταση ανθρώπινης διάσωσης, η κατάσταση εξερεύνησης των ζωνών, η τοποθεσία UAV και η αυτονομία πτήσης. Συγκρίθηκαν τα αποτελέσματα τόσο από εφαρμογές μιας διεργασίας το οποίο εκτέλεσε πολλαπλές εργασίες εντός ενός προγράμματος όσο και από πολλαπλές διεργασίες του στοχαστικού δυναμικού προγραμματισμού.

Η προσέγγιση πολλαπλών διεργασιών αποδεικνύεται αποτελεσματική για την υλοποίηση

συστημάτων που απαιτούν γρήγορη και αποδοτική λήψη αποφάσεων, καθιστώντας την ιδανική για εφαρμογές με UAV και άλλες περιοχές που αντιμετωπίζουν παρόμοιες προκλήσεις. Από τα δεδομένα που παρουσίασαν οι (Sigaud & Baffet, 2010), καταδεικνύουν την αποτελεσματικότητα της προσέγγισης πολλαπλών διεργασιών, επιτυγχάνοντας γρηγορότερη δημιουργία Πολιτικών και ενισχυμένη αυτονομία ενός συστήματος UAV. Το σύστημα ήταν πιο ανταποκρινόμενο στις απαιτήσεις και τις μεταβολές του περιβάλλοντος εκτός από την περίπτωση με 10 ζώνες, όπου η έκδοση μιας εργασίας έχει ελαφρώς καλύτερη απόδοση. Στις περισσότερες περιπτώσεις, η έκδοση πολλαπλών διεργασιών απαιτεί περισσότερους επανασχεδιασμούς από την έκδοση με μία διεργασία. Αυτό δείχνει ότι η προσέγγιση πολλαπλών διεργασιών μπορεί να εκτελέσει περισσότερες επαναλήψεις για να βελτιώσει το σχέδιο.

Η προσέγγιση πολλαπλών διεργασιών μειώνει σημαντικά τον μέγιστο χρόνο απάντησης σε σύγκριση με την προσέγγιση μόνο μίας, ιδιαίτερα εμφανής σε σενάρια με μεγαλύτερο αριθμό ζωνών. Επίσης, η ζώνη προσγείωσης ποικίλλει ανάλογα με τον αλγόριθμο και τον αριθμό των ζωνών. Φαίνεται ότι η ζώνη προσγείωσης που υπολογίζεται από τον βέλτιστο αλγόριθμο δεν ταιριάζει πάντα με τη ζώνη προσγείωσης που προκύπτει από τον αλγόριθμο που εφαρμόστηκε.

Συνολικά, η προσέγγιση πολλαπλών διεργασιών γενικά δείχνει πολλά υποσχόμενη όσον αφορά τη μείωση του μέγιστου χρόνου απάντησης, αλλά μπορεί να απαιτεί περισσότερους επανασχεδιασμούς σε σύγκριση με την προσέγγιση μόνο μίας. Η επιλογή μεταξύ υλοποιήσεων πολλαπλών εργασιών μπορεί να εξαρτάται από συγκεκριμένες απαιτήσεις απόδοσης.

Οι (Sigaud & Baffet, 2010) παρουσιάζουν τη σύγκριση μεταξύ της έκδοσης μίας διεργασίας και της έκδοσης πολλαπλών διεργασιών του αλγορίθμου στοχαστικού δυναμικού προγραμματισμού, εστιάζοντας σε διάφορα κριτήρια απόδοσης. Από την ανάλυση των κριτηρίων σύγκρισης και των αποτελεσμάτων διακρίνεται ο χαμηλότερος *Συνολικός Χρόνος Βελτιστοποίησης*. Αναφέρεται στο άθροισμα των χρόνων βελτιστοποίησης τόσο για τον αρχικό σχεδιασμό όσο και για όλες τις διαδικασίες επανασχεδιασμού.

Η εφαρμογή μέσω πολλαπλών διεργασιών αναμένεται να έχει *Μικρότερο Συνολικό Χρόνο Βελτιστοποίησης* σε σύγκριση με την έκδοση μίας διεργασίας λόγω των δυνατοτήτων παράλληλης επεξεργασίας. Ενώ ο *Αριθμός Ακολουθιών Τοπικού Ανασχεδιασμού* υποδεικνύει πόσες ακολουθίες επανασχεδιασμού εκτελούνται κατά τη διάρκεια της αποστολής. Επιπλέον, ενδέχεται να απαιτεί λιγότερες ακολουθίες επανασχεδιασμού λόγω της ικανότητάς της να βελτιστοποιεί συνεχώς στο

παρασκήνιο.

Ο μέγιστος χρόνος που απαιτείται για την αποστολή μιας ενέργειας στον επόπτη μετά τη λήψη της τρέχουσας κατάστασης αναμένεται να έχει μεγαλύτερη ανταπόκριση σε σύγκριση με την έκδοση με μίας διεργασίας, εξασφαλίζοντας ταχύτερη λήψη αποφάσεων. Και οι δύο εκδόσεις αναμένεται να παρέχουν παρόμοιες ζώνες προσγείωσης, καθώς χρησιμοποιούν τον ίδιο αλγόριθμο βελτιστοποίησης.

Οι δοκιμές διεξήχθησαν σε έναν ενσωματωμένο επεξεργαστή επεξεργασίας εικόνας και λήψης αποφάσεων, με ενέργειες που διαρκούσαν περίπου 50 δευτερόλεπτα κατά μέσο όρο. Σημειώνεται ότι λόγω της μείωσης της μεταβλητής αυτονομίας πτήσης κατά τον επανασχεδιασμό, μπορεί να απαιτηθεί σημαντικός αριθμός ακολουθιών επανασχεδιασμού στην έκδοση μίας διεργασίας πριν από την εφαρμογή μίας ενέργειας, κάτι που δεν συμβαίνει με την υλοποίηση πολλαπλών διεργασιών ανά πάσα στιγμή.

Η χρήση πολλαπλών διεργασιών οποτεδήποτε τονίζεται ως ικανή να παράγει εφαρμόσιμες και σχετικές πολιτικές σε πολύ σύντομο χρονικό διάστημα, αποδεικνύοντας την αποτελεσματικότητά της σε σενάρια λήψης αποφάσεων σε πραγματικό χρόνο. Αυτή η προσέγγιση αξιοποιεί την παράλληλη επεξεργασία για να επιταχύνει τις διαδικασίες βελτιστοποίησης και λήψης αποφάσεων, διασφαλίζοντας έγκαιρες αποκρίσεις ακόμη και σε δυναμικά περιβάλλοντα.

ΚΕΦΑΛΑΙΟ 6

Επίλογος-Συμπεράσματα

Τα Μη Επανδρωμένα Συστήματα έχουν κερδίσει σημαντική προσοχή τα τελευταία χρόνια λόγω της πιθανής επίδρασής τους σε διάφορους τομείς όπως η επιτήρηση, η στρατιωτική χρήση, επιχειρήσεις έρευνας και διάσωσης κλπ. Μια κρίσιμη πτυχή των Μη Επανδρωμένων Συστημάτων είναι η ικανότητά τους να λαμβάνουν αποφάσεις σε πραγματικό χρόνο με βάση το περιβάλλον. Οι αλγόριθμοι σχεδιασμού της Διαδικασίας Απόφασης Markov (MDP) έχουν αναδειχθεί ως μια πολλά υποσχόμενη προσέγγιση για να καταστεί δυνατή η αυτόνομη λήψη αποφάσεων. Αυτή η εργασία στοχεύει να αναδείξει την εφαρμογή των αλγορίθμων σχεδιασμού MDP για τη λήψη αποφάσεων σε πραγματικό χρόνο επί των Μη Επανδρωμένων Συστημάτων για την εύρεση ασφαλούς ζώνης προσγείωσης.

Για την αξιολόγηση της αποτελεσματικότητας της μεθοδολογίας, παρουσιάστηκε μια περιπτωσιολογική βιβλιογραφική μελέτη που περιελάβανε πραγματικό σενάριο. Αυτή η μελέτη περιλαμβάνει διάφορες εφαρμογές Μη Επανδρωμένων Συστημάτων, όπως για την εύρεση ασφαλούς ζώνης προσγείωσης όπως προαναφέρθηκε. Τα αποτελέσματα που προκύπτουν από αυτές τις μελέτες χρησιμοποιούνται για την αξιολόγηση της απόδοσης των αλγορίθμων σχεδιασμού Μαρκοβιανών Διαδικασιών Αποφάσεων όσον αφορά την ακρίβεια απόφασης, την υπολογιστική απόδοση και την προσαρμοστικότητα σε δυναμικά περιβάλλοντα.

Τα δυνατά και τα αδύνατα σημεία της περιπτωσιολογικής μελέτης διερευνώνται στο πλαίσιο της εφαρμογής τους σε Μη Επανδρωμένα Συστήματα. Επιπλέον, η καταλληλόλητα αυτών των αλγορίθμων για λήψη αποφάσεων σε πραγματικό χρόνο αξιολογείται με βάση την υπολογιστική πολυπλοκότητά τους και την ικανότητά τους να χειρίζονται άγνωστα περιβάλλοντα.

Τα αποτελέσματα που προέκυψαν από τη βιβλιογραφική περιπτωσιολογική μελέτη (Sigaud & Baffet, 2010) δείχνει ότι οι αλγόριθμοι σχεδιασμού Μαρκοβιανών Διαδικασιών Αποφάσεων είναι ικανοί να επιτύχουν σχεδόν βέλτιστη λήψη αποφάσεων σε σενάρια πραγματικού χρόνου.

Οι αλγόριθμοι επανάληψης τιμών και επανάληψης πολιτικής επιδεικνύουν υψηλή ακρίβεια, αλλά η υπολογιστική πολυπλοκότητά τους μπορεί να περιορίσει την εφαρμογή τους σε Μη Επανδρωμένα Συστήματα με περιορισμούς πόρων. Με βάση την ανάλυση διάφορων αλγορίθμων σχεδιασμού Μαρκοβιανών Διαδικασιών Αποφάσεων και την εφαρμογή τους στη λήψη αποφάσεων σε πραγματικό χρόνο επί των Μη Επανδρωμένων Συστημάτων, μπορεί να αναδειχθεί το συμπέρασμα ότι οι αλγόριθμοι σχεδιασμού Μαρκοβιανών Διαδικασιών Αποφάσεων παρουσιάζουν μεγάλες δυνατότητες να επιτρέπουν την αυτόνομη λήψη αποφάσεων σε Μη Επανδρωμένα Συστήματα.

Ωστόσο, η επιλογή του αλγορίθμου θα πρέπει να εξετάζεται προσεκτικά με βάση τις συγκεκριμένες απαιτήσεις της εφαρμογής Μη Επανδρωμένων Συστημάτων, όπως υπολογιστικούς πόρους, δυναμική περιβάλλοντος και ο παράγοντας αβεβαιότητας. Απαιτείται περαιτέρω έρευνα για την αντιμετώπιση των προκλήσεων που έχουν εντοπιστεί και τη βελτιστοποίηση της απόδοσης των συστημάτων λήψης αποφάσεων στα Μη Επανδρωμένα Συστήματα.

Για να βελτιωθεί η εφαρμογή των αλγορίθμων σχεδιασμού Μαρκοβιανών Διαδικασιών Αποφάσεων για τη λήψη αποφάσεων σε Μη Επανδρωμένα Συστήματα, η μελλοντική έρευνα θα πρέπει να επικεντρωθεί στην αντιμετώπιση των περιορισμών και των προκλήσεων που προσδιορίζονται σε αυτή την βιβλιογραφική εργασία.

Αυτό περιλαμβάνει την ανάπτυξη αποτελεσματικών ενεργειών για χώρους συνεχούς κατάστασης και ενέργειας, την ενσωμάτωση δεδομένων στους αισθητήρες σε πραγματικό χρόνο και την εξέταση σε πιο περίπλοκα δυναμικά περιβάλλοντα. Η συνεργασία μεταξύ ερευνητών από τους τομείς της επιστήμης και της τεχνολογίας είναι απαραίτητη για την προώθηση της καινοτομίας και τη διασφάλιση της ευρείας υιοθέτησης των αλγορίθμων σχεδιασμού Μαρκοβιανών Διαδικασιών Αποφάσεων σε αυτόνομα εναέρια οχήματα.

Βιβλιογραφία

Bavakutty M., Salih Muhammed T. K, and Haneefa K. Mohamed (2006), Research on library computerization. New Delhi: Ess Ess.

Bishop M. Christopher (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2.

Chhaya Ku. A. Khanzode & Dr. Ravindra D. Sarode (2020), Advantages And Disadvantages Of Artificial Intelligence And Machine Learning: *A Literature Review,* 2277-3584.

Fabiani P., Fuertes V., Le Besnerais G., Mampey r., Piquereau A. and Teichteil F., (2007) *"The ReSSAC autonomous helicopter: Flying in a non-cooperative uncertain world with embedded vision and decision making"*, A.H.S. Forum

Howard R.A. (1960), Dynamic Programming and Markov Processes, The Technology Press of The Massachusetts Institute of Technology and John Wiley & Sons, Inc., New York, London.

Norris J. R. (1997), *Markov Chains,* University of Cambridge, New York, NY 10013-2473, USA

Sigaud O. & Buffet O. (2010), Markov Decision Processes in Artificial Intelligence, *MDPs,* Beyond MDPs and Applications, Wiley.

Sutton R.S. and Barto A.G. (2014), Reinforcement Learning: *An Introduction* Second edition, in progress, The MIT Press Cambridge, Massachusetts London, England.

Poole D., Macworth A., & Goebel, R. (1998). Computational Intelligence: A Logical Approach. New York: Oxford University Press.

Puterman M. L. (1994), Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley.