Βέλτιστη Τοπολογία Δικτύων Διανομής Ηλεκτρικής Ενέργειας με Εφαρμογή της Ενισχυτικής Μηχανικής Μάθησης

Ι. Γ. ΒΛΑΧΟΓΙΑΝΝΗΣ
Δρ Ηλεκτρολόγος Μηχανικός Α.Π.Θ.

Ν. Δ. ΧΑΤΖΗΑΡΓΥΡΙΟΥ
Καθηγητής Ε.Μ.Π.

Περιληψη
Η εργασία παρουσιάζει τη μέθοδο Ενισχυτικής Μηχανικής Μάθησης (ΕΜ) με σκοπό τη βέλτιστη τοπολογία των δικτύων διανομής ηλεκτρικής ενέργειας (ΔΔΗΕ). Η βέλτιστη τοπολογία αφορά στην επιλογή του κατάλληλου συνόλου των κλάδων που πρόκειται να απενεργοποιηθούν, ένας από κάθε βρόχο, έτσι ώστε το τελικό ΔΔΗΕ να έχει τη βέλτιστη απόδοση. Κριτήριο για τη βέλτιστη απόδοση θεωρείται η ελαχιστοποίηση των απωλειών ενεργού ισχύος ενώ ταυτόχρονα πρέπει να ικανοποιούνται τα όρια των τάσεων. Η μέθοδος ΕΜ αντιμετωπίζει τη βέλτιστη τοπολογία των ΔΔΗΕ ως πρόβλημα λήψης απόφασης πολλών επιπέδων απεικονίζοντας εμπειρικά καταστάσεις λειτουργίας του ΔΔΗΕ σε συγκεκριμένες δράσεις μέσω βαθμών επιβράβευσης. Ο αλγόριθμος αυτός πειραματικά εφαρμόζεται στη βέλτιστη τοπολογία ενός ΔΔΗΕ εφαρμογής 33 ζυγών. Τα αποτελέσματα που προκύπτουν συγκρίνονται με εκείνα άλλων εξελικτικών μεθόδων τεχνητής νοημοσύνης (ΤΝ).

1.ΕΙΣΑΓΩΓΗ

Η διαμόρφωση ενός δικτύου διανομής ηλεκτρικής ενέργειας (ΔΔΗΕ) στοχεύει στη βέλτιστη λειτουργία του ικανοποιώντας ταυτόχρονα τους φυσικούς και λειτουργικούς περιορισμούς [10].

Ένα από τα κριτήρια βέλτιστης λειτουργίας είναι η ελαχιστοποίηση των απωλειών ενεργού ισχύος ενώ ταυτόχρονα πρέπει να ικανοποιούνται τα όρια των τάσεων λειτουργίας. Για τη λύση του συγκεκριμένου προβλήματος έχουν αναπτυχθεί πολλοί αλγόριθμοι, οι οποίοι βασίζονται σε εξελικτικές υπολογιστικές τεχνικές [1-5].

Ωστόσο, αυτές οι μέθοδοι δεν μπορούν να παρέχουν τις βέλτιστες τοπολογίες συνολικά για ολόκληρη τη λειτουργική περίοδο του ΔΔΗΕ.

Σε αυτή την εργασία, το πρόβλημα της βέλτιστης τοπολογίας των ΔΔΗΕ επιλύεται μέσω της Ενισχυτικής Μηχανικής Μάθησης (ΕΜ) [6-9]. Η ΕΜ προέρχεται από τη θεωρία του ελέγχου και του δυναμικού προγραμματισμού και στόχο έχει την προσέγγιση εμπειρικών λύσεων σε προβλήματα άγνωστης δυναμικής [8]. Θεωρητικά, έχουν πραγματοποιηθεί αποφασιστικά βήματα με σκοπό τη γρήγορη σύγκληση της ΕΜ και την εφαρμογή της σε μη γραμμικά συστήματα [6, 8] με την ανάπτυξη πολύ αποδοτικών αλγορίθμων. Τελευταία η ραγδαία ανάπτυξη των υπολογιστικών συστημάτων βοήθησε πολύ στην εξαιρετική απόδοση των αλγορίθμων ΕΜ [6, 8]. Για την εφαρμογή του προτεινόμενου αλγορίθμου ΕΜ, το πρόβλημα της βέλτιστης τοπολογίας των ΔΔΗΕ μετατρέπεται σε πρόβλημα λήψης απόφασης πολλών επιπέδων.

Οι βέλτιστες δράσεις (από κάθε ενδεχόμενο βρόχο του ΔΔΗΕ αποκόπτεται και ένας κλάδος) απεικονίζονται εμπειρικά στις καταστάσεις λειτουργίας του ΔΔΗΕ. Οι δράσεις βασίζονται σε επιβραβεύσεις, οι οποίες εκφράζουν την επιτυχία αυτών των δράσεων σε όλη τη λειτουργική περίοδο. Κριτήριο της επιβράβευσης αποτελεί η ελαχιστοποίηση των απωλειών ενεργού ισχύος. Επίσης, πρέπει να τηρούνται τα όρια λειτουργίας των τάσεων.

Στην εργασία ο αλγόριθμος εκμάθησης Q [6] προσαρμόζεται για τη βέλτιστη διαμόρφωση των ΔΔΗΕ, ωστόσο, ο αλγόριθμος είναι γενικός και μπορεί να εφαρμοστεί σε μεγάλη ομάδα προβλημάτων βελτιστοποίησης των συστημάτων ηλεκτρικής ενέργειας. Η εργασία στη συνέχεια οργανώνεται σε τέσσερις ενότητες.

Η Ενότητα 1 περιγράφει τη μέθοδο της Ενισχυτικής Μηχανικής Μάθησης (ΕΜ).

Στην Ενότητα 2, ο αλγόριθμος μάθησης Q [6] προσαρμόζεται για τη βέλτιστη τοπολογία των ΔΔΗΕ.

Στην Ενότητα 3, παρουσιάζονται τα αποτελέσματα που προκύπτουν από την εφαρμογή του αλγόριθμου σε ΔΔΗΕ εφαρμογής 33 ζυγών.

Τα αποτελέσματα συγκρίνονται με εκείνα ενός εξελικτικού αλγορίθμου Τεχνητής Νοημοσύνης (ΤΝ) [1], καταδεικνύοντας την υπεροχή της ΕΜ. Επιπλέον, γίνεται εμφανές το πλεονέκτημα του προτεινόμενου αλγορίθμου καθώς παρέχει τη βέλτιστη τοπολογία για ολόκληρη τη λειτουργική περίοδο των ΔΔΗΕ.

Τέλος, στην Ενότητα 4 εξάγονται γενικά συμπεράσματα της εφαρμογής.

2. ΕΝΙΣΧΥΤΙΚΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ (ΕΜ)

Οι τεχνικές ΕΜ είναι απλοί επαναληπτικοί αλγόριθμοι, οι οποίοι μαθαίνουν εμπειρικά να ενεργούν κατά βέλτιστο τρόπο μέσω της εξερεύνησης ενός συστήματος άγνωστου δυναμικής [6-10]. Η ΕΜ υποθέτει ότι ο «κόσμος» μπορεί να περιγραφεί με την χρήση ενός συνόλου S από καταστάσεις (states) και ένας «πράκτορας» ("agent") μπορεί να επιλέγει μία δράση από ένα σύνολο δράσεων A (actions). Η λειτουργία εκμάθησης του ΕΜ υλοποιείται σε διακριτά βήματα. Σε κάθε βήμα μάθησης ο πράκτορας εξετάζει την παρούσα κατάσταση s του «κόσμου» ($s \in S$), και επιλέγει μία δράση $a \in A$, η οποία μεγιστοποιεί μακροπρόθεσμα το αναμενόμενο κέρδος επιβράβευσης [6-8]. Στη συνέχεια, αφού εκτελέσει τη δράση (a), δίδεται στον πράκτορα μια άμεση επιβράβευση (reward) $r \in R$, που εκφράζει την αποτελεσματικότητα της δράσης που επιλέχθηκε παρατηρώντας το αποτέλεσμά της στη νέα κατάσταση $s' \in S$ του «κόσμου». Στην εργασία χρησιμοποιήθηκε ο ΕΜ αλγόριθμος μάθησης Q [6]. Στην ενισχυτική μάθηση Q η βέλτιστη συνάρτηση επιβράβευσης ορίζεται με χρήση της εξίσωσης Bellman, ως εξής:

$$Q^*(s,a) = E\left(r(s,a) + \gamma \max_{a'} Q^*(s',a')\right) \qquad (2.1)$$

Η εξίσωση αυτή αναπαριστά το αναμενόμενο άθροισμα των επιβραβεύσεων που λαμβάνεται ξεκινώντας από μια αρχική κατάσταση (s), εκτελώντας τη δράση (a) και επιλέγοντας βέλτιστες δράσεις (a′) στις επόμενες αναζητήσεις. Αυτό γίνεται σε περιορισμένο ή θεωρητικά άπειρο χρονικά ορίζοντα μέχρι να επιτευχθεί η βέλτιστη τιμή της συνάρτησης Q (Q*(s, a)). Η εκπτωτική παράμετρος γ ($0 \le \gamma \le 1$) χρησιμοποιείται για να μειωθεί εκθετικά το βάρος των επιβραβεύσεων που λαμβάνεται κατά τις επόμενες αναζητήσεις [6-8]. Εφόσον έχουμε τη βέλτιστη τιμή Q*(s, a) είναι εύκολο να καθοριστεί η βέλτιστη δράση a* με τη χρήση της βέλτιστης πολιτικής [6-9]. Ένας απλός τρόπος είναι να ερευνηθούν όλες οι δυνατές δράσεις (a) για μία δεδομένη κατάσταση (s) και να επιλεγεί εκείνη με τη μεγαλύτερη τιμή:

$$a^* = \arg\max_a Q^*(s,a) \qquad (2.2)$$

Η συνάρτηση Q (μνήμη-Q) συνήθως αποθηκεύεται σε κάποιον πίνακα με διαστάσεις τον αριθμό των καταστάσεων και τον αριθμό των δράσεων. Λαμβάνοντας αρχικά αυθαίρετες τιμές μπορεί, μέσω επαναλήψεων, να προσεγγιστεί η βέλτιστη συνάρτηση Q σύμφωνα με τα προκαθορισμένα κριτήρια.

Στη συνέχεια η εγγραφή στον πίνακα που απεικονίζει την απόδοση της δράσης (a) πάνω στην κατάσταση (s) γίνεται σύμφωνα με την παρακάτω επαναληπτική εξίσωση [6]:

$$Q(s,a) = (1-\alpha)\cdot Q(s,a) + \alpha\left(r + \gamma \max_{a'} Q(s',a')\right) \qquad (2.3)$$

Είναι σημαντικό να σημειωθεί ότι η νέα τιμή της μνήμης Q(s,a) βασίζεται τόσο στην τρέχουσα τιμή της Q(s,a), (μνήμη-Q) όσο και στις τιμές (άμεσες επιβραβεύσεις) των δράσεων που λαμβάνονται από τις επόμενες δράσεις. Έτσι, η παράμετρος a ($0 \le a \le 1$) αναπαριστάνει το ποσοστό της νέας γνώσης που εναποτίθεται στη μνήμη επηρεάζοντας έτσι τον αριθμό των επαναλήψεων μάθησης. Η παράμετρος (1-a) εκφράζει το συνολικό ποσοστό των τιμών Q που παραμένουν ως «μνήμη» στη συνάρτηση Q [6].

3. ΕΦΑΡΜΟΓΗ ΤΗΣ ΕΜ ΣΤΗ ΒΕΛΤΙΣΤΗ ΤΟΠΟΛΟΓΙΑ ΤΩΝ ΔΔΗΕ

Για τους σκοπούς της συγκεκριμένης μελέτης θεωρούμε μία δυαδική ταξινόμηση του «κόσμου» των ΔΔΗΕ. Δηλαδή ο «κόσμος» των καταστάσεων λύσης ($s \in S$) από την ΕΜ αποτελείται από αποδεκτά λειτουργικά «σημεία», όπου όλοι οι περιορισμοί ικανοποιούνται, και μη αποδεκτά, όπου παραβιάζεται οποιοσδήποτε από τους λειτουργικούς περιορισμούς. Το διάνυσμα των δράσεων ($a \in A$) είναι το σύνολο από τις πιθανές δράσεις (αφαίρεση ενός κλάδου από κάθε εν δυνάμει βρόχο του ΔΔΗΕ). Ο αλγόριθμος εξελίσσεται ως εξής:

Επιλέγεται ένα τυχαίο λειτουργικό σημείο που συμπεριλαμβάνει μια τυχαία κατανομή φορτίου καθώς και ένα σύνολο δράσεων. Ο πράκτορας παρατηρεί την κατάσταση (s) του συστήματος, όπως αυτή προκύπτει από τη λύση της ροής φορτίου, και στη συνέχεια επιλέγει ένα συνδυασμό δράσεων (a) από το σύνολο των πιθανών δράσεων. Εκτελείται μία νέα ροή φορτίου. Ο πράκτορας παρατηρεί τη νέα κατάσταση που προκύπτει από τη λύση (s') λαμβάνοντας μία άμεση επιβράβευση (r): $SxA \rightarrow R$, η οποία εκφράζει τις απώλειες ενεργού ισχύος του ΔΔΗΕ.

Στη συνέχεια επιλέγεται ένας νέος συνδυασμός δράσεων, οδηγώντας σε νέα λύση της ροής φορτίου και νέα επιβράβευση.

Η επιλογή νέων δράσεων επαναλαμβάνεται μέχρι να μην παρουσιάζονται πλέον αλλαγές στην τιμή της άμεσης επιβράβευσης ή την επιλεγόμενη δράση. Στόχος του πράκτορα είναι να προσδιορίσει την βέλτιστη συνάρτηση Q (Q*(s,a)) χρησιμοποιώντας τις αντιστοιχίες των καταστάσεων προς τις δράσεις ($S \rightarrow A$), ώστε να μεγιστοποιηθούν μακροπρόθεσμα οι επιβραβεύσεις. Η διαδικασία επαναλαμβάνεται για μεγάλο αριθμό λειτουργικών σημείων που καλύπτουν ολόκληρη τη λειτουργική περίοδο του ΔΔΗΕ. Ο πράκτορας βρίσκει τις βέλτιστες δράσεις (a*) χρησιμοποιώντας τη βέλτιστη πολιτική που περιγράφεται από τη σχέση 2.3. Ο Πίνακας 1 παρουσιάζει τον αλγόριθμο ΕΜ στη βέλτιστη τοπολογία των ΔΔΗΕ.

3.1. Διανύσματα κατάστασης

Οι καταστάσεις λειτουργίας (s) ενός συστήματος ΔΔΗΕ διακρίνονται ως εξής:

Όταν μία από τις μεταβλητές (εδώ τιμές των τάσεων) βρίσκεται εκτός των ορίων λειτουργίας της, θωρείται ότι η κατάσταση είναι επιπέδου –1, διαφορετικά θεωρείται κατάσταση επιπέδου 0. Κατά συνέπεια, αν έχουμε n λειτουργικές μεταβλητές, ο συνολικός αριθμός των δυνατών καταστάσεων είναι:

$$\overline{\overline{S}} = 2^n \qquad (3.1)$$

Στην παρούσα εφαρμογή η χαμηλότερη τάση σε κάθε βρόχο πρέπει να βρίσκεται πάνω από τα κατώτατα όρια και αντίστοιχα η υψηλότερη κάτω από τα ανώτατα.

Πίνακας 1: Εφαρμογή του αλγορίθμου ΕΜ στη βέλτιστη διαμόρφωση των ΔΔΗΕ.

1. Αρχικοποίησε τη μνήμη Q(s,a)=0.0 και τις άμεσες επιβραβεύσεις r(s,a)=0.0, ∀ s ∈ S, ∀ a ∈ A
2. Επανάλαβε για τυχαίο αριθμό λειτουργικών σημείων από όλη την λειτουργική περίοδο του ΔΔΗΕ (τυχαία διακύμανση φορτίου)
2.1. Επανάλαβε…
2.1.1. Παρατήρησε την κατάσταση (s) της λύσης της ροής φορτίου
2.1.2. Επίλεξε ένα διάνυσμα δράσεων (a)
2.1.3. Εκτέλεσε τη ροή φορτίου
2.1.4. Παρατήρησε την νέα κατάσταση (s') που προκύπτει από τη ροή φορτίου και υπολόγισε την άμεση επιβράβευση (σχ. 2.3)
2.1.5. Ενημέρωσε τη συνάρτηση Q (σχ. 1.3)
2.1.6. Αντικατάστησε την παλιά με τη νέα κατάσταση (s←s')
…μέχρι τον προσδιορισμό της βέλτιστης συνάρτησης Q (καμία περαιτέρω αλλαγή στην επιβράβευση ή στο διάνυσμα των επιλεγόμενων δράσεων)

3.2. Διανύσματα δράσεων

Αν κάθε δράση (a) διακρίνεται σε d_{u_i} επίπεδα (αριθμός κλάδων που μπορούν να απενεργοποιηθούν από κάθε βρόχο του ΔΔΗΕ), ο συνολικός αριθμός των δράσεων είναι:

$$\overline{\overline{A}} = \prod_{i=1}^{m} d_{u_i} \qquad (3.2)$$

Όπου το m εκφράζει το συνολικό αριθμό μεταβλητών ελέγχου (αριθμός των βρόχων).

3.3. Επιβραβεύσεις (r)

Η βέλτιστη διαμόρφωση προϋποθέτει την επιλογή του καλύτερου συνδυασμού κλάδων που θα απενεργοποιηθούν, έναν από κάθε εν δυνάμει βρόχο, έτσι ώστε το ΔΔΗΕ που προκύπτει να έχει τη βέλτιστη (επιθυμητή) απόδοση. Ανάμεσα σε πολλά κριτήρια βέλτιστης απόδοσης ενός ΔΔΗΕ, επιλέχθηκε η ελαχιστοποίηση των απωλειών ενεργού ισχύος. Η εφαρμογή του αλγόριθμου ΕΜ στη βέλτιστη τοπολογία των ΔΔΗΕ είναι συνδεδεμένη με την λήψη της άμεσης επιβράβευσης (r), τέτοιας ώστε η επαναληπτική τιμή της συνάρτησης Q (σχ. 2.3) να μεγιστοποιείται ενώ ταυτόχρονα να ικανοποιείται η ελαχιστοποίηση των συνολικών απωλειών ενεργού ισχύος καθ' όλη τη λειτουργική περίοδο του ΑΔΔΗΕ. Επομένως η άμεση επιβράβευση (r) υπολογίζεται ως εξής:

r = -Ολικές Απώλειες Ενεργού Ισχύος (3.3)

4. ΑΠΟΤΕΛΕΣΜΑΤΑ

Ο προτεινόμενος αλγόριθμος (Πίνακας 1) εφαρμόζεται για τη βέλτιστη τοπολογία του ΔΔΗΕ εφαρμογής 33 ζυγών. Το μονογραμμικό διάγραμμα και τα δεδομένα του συγκεκριμένου δικτύου παρατίθενται στο παράρτημα καθώς και στην [1]. Οι δράσεις αποτελούν τα σύνολα των κλάδων που θα απενεργοποιηθούν, ένα από κάθε βρόχο. Υπάρχουν πέντε εν δυνάμει βρόχοι και κατά συνέπεια κάθε διάνυσμα δράσεων της ΕΜ είναι 1x5. Στον Πίνακα 2 εμφανίζονται οι κλάδοι που αποτελούν κάθε βρόχο. Σύμφωνα με τον Πίνακα 2, ο συνολικός αριθμός πιθανών δράσεων υπολογίζεται σε 10x7x7x16x11 = 86240. Εφόσον τα κατώτερα και ανώτερα μεγέθη τάσης (v) σε κάθε βρόχο περιορίζονται στα λειτουργικά όρια [0.96, 1.05] pu, ο συνολικός αριθμός των πιθανών καταστάσεων λύσης της ΕΜ υπολογίζεται σε 2^5=32.

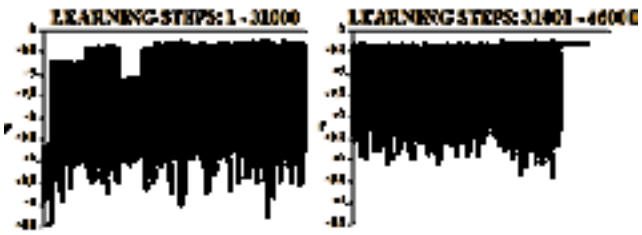
Πίνακας 2: Κλάδοι που θα απενεργοποιηθούν σε κάθε βρόχο του ΔΔΗΕ 33 ζυγών.

Βρόχος	Κλάδοι
1	2-3-4-5-6-7-18-19-20-33
2	9-10-11-12-13-14-34
3	8-9-10-11-21-33-35
4	6-7-8-15-16-17-25-26-27-28-29-30-31-32-34-36
5	3-4-5-22-23-24-25-26-27-28-37

Ο αλγόριθμος ΕΜ (Πίνακας 1) μπορεί να εφαρμοστεί σε μεγάλο πλήθος από συνδυασμούς φορτίων (λειτουργικά σημεία) που έχουν επιλεγεί από ολόκληρη την λειτουργική περίοδο του ΔΔΗΕ. Αρχικά εφαρμόζεται σε συγκεκριμένη κατανομή φορτίου που αντιστοιχεί στη μέση τιμή σε κάθε κόμβο κατανάλωσης [1]. Στην περίπτωση αυτή οι παράμετροι της μάθησης του ΕΜ επιλέγονται α=0.99 και γ=0.01. Το Σχήμα 1 εμφανίζει την άμεση επιβράβευση r (σχ. 2.3) που λήφθηκε σε κάθε βήμα της μάθησης. Κάθε βήμα της μάθησης αντιστοιχεί σε μία επανάληψη του αλγόριθμου μάθησης ΕΜ (Πίνακας 1).

Ο πράκτορας εκτέλεσε περίπου 46000 βήματα μάθησης

για να προσδιορίσει τις βέλτιστες δράσεις. Συνολικά απαιτήθηκαν 100 sec σε PC 1.4 GHz Pentium-IV για τη σύγκληση της ΕΜ. Επίσης στο Σχήμα 1 φαίνεται ότι η σύγκληση του αλγόριθμου μάθησης πέτυχε μέγιστη τιμή επιβράβευσης -0.354, αντιστοιχώντας το βέλτιστο συνδυασμό δράσεων στην πιο επιθυμητή κατάσταση λύσης της ΕΜ.



Σχήμα 1. Άμεσες επιβραβεύσεις του αλγορίθμου μάθησης ΕΜ. Figure 1:(Immediate rewards of Q-learning algorithm)

Ο Πίνακας 3 εμφανίζει το βέλτιστο συνδυασμό δράσεων (το καλύτερο σύνολο ανενεργών κλάδων) 7-10-13-31-25 και τις απώλειες ενεργού ισχύος που υπολογίζονται στα 110,05 kW. Επιπλέον, στον Πίνακα 3 εμφανίζονται οι τάσεις που επιτυγχάνονται από τον εξελικτικό αλγόριθμο TN [1]. Σύμφωνα με αυτόν, καλύτερος συνδυασμός δράσεων θεωρείται η αποκοπή των κλάδων 6-14-9-32-37.

Ο Πίνακας 3 παρέχει επίσης τις τάσεις του ΔΔΗΕ εφαρμογής των 33 ζυγών για τη βασική περίπτωση με ανενεργούς κλάδους τους 33-34-35-36-37. Συγκρίνοντας τη βέλτιστη λύση της ΕΜ με την αντίστοιχη του εξελικτικού αλγορίθμου TN [1], η πρώτη υπερέχει καθώς όλοι οι περιορισμοί τάσης ικανοποιούνται και οι απώλειες ενεργού ισχύος είναι μικρότερες (110.05 kW έναντι 118.37 kW).

Ο αλγόριθμος ΕΜ προσφέρει επίσης, on-line έλεγχο του ΔΔΗΕ σε τυχαίο δυναμικό περιβάλλον [8]. Μια τέτοια περίπτωση θεωρείται όταν το φορτίο του συστήματος κυμαίνεται τυχαία εντός μιας καθορισμένης λειτουργικής περιόδου. Η διακύμανση του φορτίου εδώ θεωρείται προσεγγιστικά κυκλική με περίοδο 50 βημάτων μάθησης ΕΜ και διαμορφώνεται σύμφωνα με την εξίσωση:

$$z(ls) = z_{max} \cdot \sin\left(\frac{2 \cdot \pi \cdot ls}{50}\right) \qquad (4.1)$$

όπου z_{max} συμβολίζει το μέγιστο φορτίο σε κάθε ζυγό.

Στη συγκεκριμένη περίπτωση θέτουμε τις εξής παραμέτρους α=0.1 και γ=0.98.

Η γνώση που αποκτά σταδιακά ο πράκτορας σε όλη τη διάρκεια της λειτουργικής περιόδου απεικονίζεται στο Σχήμα 2. Η σύγκληση του αλγορίθμου ΕΜ χρειάστηκε περίπου 78.000 βήματα μάθησης.

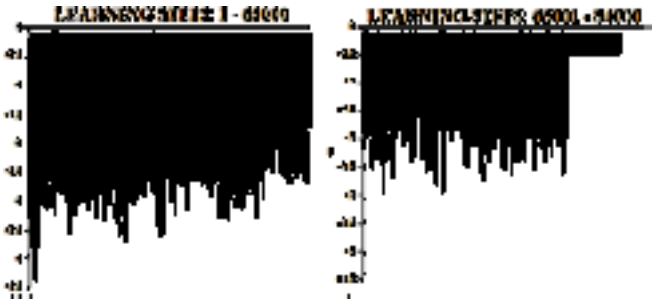
Κάθε βήμα μάθησης αντιστοιχεί σε μία επανάληψη του αλγορίθμου μάθησης (Πίνακας 1). Η συνολική διάρκεια υπολογισμού ήταν 190 sec σε PC 1.4 GHz Pentium-IV. Στο Σχήμα 2 επίσης απεικονίζεται η σύγκληση του αλγορίθμου μάθησης ΕΜ σε ένα μέγιστο εύρος επιβραβεύσεων μεταξύ -0.467 και -0.155 p.u. καθ' όλη τη λειτουργική περίοδο του ΔΔΗΕ, αντιστοιχώντας έτσι ένα βέλτιστο συνδυασμό δράσεων στις καλύτερες (επιθυμητές) καταστάσεις λύσης της ΕΜ.

Πίνακας 3: Συνολικές απώλειες ενεργού ισχύος και τάσεις ζυγών από την ΕΜ, τον εξελικτικό αλγόριθμο TN και την αρχική κατάσταση του ΔΔΗΕ 33 ζυγών.

Μέθοδος	ΕΜ	Εξελικτική μέθοδος TN [1]	Αρχική κατάσταση
Βέλτιστη δράση	6-13-10-31-25	7-14-9-32-37	33-34-35-36-37
Ολικές απώλειες ενεργού ισχύος	110.05kW	139.83kW	181.53kW
Ζυγός		*Τάσεις*	
1	1.020	1.020	1.020
2	1.017	1.017	1.017
3	1.007	1.007	1.003
4	1.005	1.003	0.996
5	1.003	1.000	0.990
6	0.998	0.991	0.974
7	0.998	0.982	0.972
8	0.986	0.984	0.961
9	0.981	0.981	0.956*
10	0.982	0.982	0.952*
11	0.986	0.983	0.952*
12	0.987	0.984	0.951*
13	0.985	0.980	0.946*
14	0.976	0.979	0.945*
15	0.977	0.979	0.944*
16	0.974	0.977	0.944*
17	0.969	0.973	0.943*
18	0.968	0.972	0.944*
19	1.015	1.015	1.016
20	0.997	0.995	1.012
21	0.992	0.990	1.010
22	0.988	0.986	1.009
23	1.001	1.004	0.999
24	0.989	0.997	0.990
25	0.980	0.993	0.985
26	0.982	0.989	0.972
27	0.982	0.987	0.970
28	0.980	0.978	0.960
29	0.978	0.971	0.953*
30	0.976	0.968	0.949*
31	0.974	0.965	0.944*
32	0.966	0.962	0.943*
33	0.967	0.971	0.942*

* Η τάση παραβιάζει το κατώτατο όριο

Η τελική βέλτιστη δράση όπως προκύπτει από τη βέλτιστη πολιτική (σχ. 2.3) υποδεικνύει την απενεργοποίηση των κλάδων 6-10-8-32-37 από το ΔΔΗΕ εφαρμογής 33-ζυγών, ελαχιστοποιώντας τις συνολικές απώλειες ενεργού ισχύος και ικανοποιώντας τους περιορισμούς στα όρια των τάσεων καθ' όλη τη διάρκεια της λειτουργικής περιόδου.



Σχήμα 2: Άμεσες επιβραβεύσεις του αλγορίθμου ΕΜ με κυμαινόμενο φορτίο.

Figure 2: (Immediate rewards of Q-learning algorithm over the whole planning period)

5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην εργασία αυτή εφαρμόστηκε η μέθοδος ΕΜ στην βέλτιστη τοπολογία των ΔΔΗΕ. Εφαρμόζεται ένας επαναληπτικός αλγόριθμος ΕΜ με σκοπό να παρέχει το βέλτιστο συνδυασμό δράσεων (το σύνολο των κλάδων που πρόκειται να απενεργοποιηθούν, ένα από κάθε εν' δυνάμει βρόχο του ΔΔΗΕ) ικανοποιώντας ταυτόχρονα τα όρια λειτουργίας των δεσμευμένων μεταβλητών (τάσεις) και ελαχιστοποιώντας τις συνολικές απώλειες ενεργού ισχύος. Οι δράσεις αντιστοιχίζονται με εμπειρικούς (μαθησιακούς) κανόνες στις καταστάσεις λειτουργίας μέσω βαθμών επιβράβευσης. Ως συνάρτηση επιβράβευσης ορίζονται οι συνολικές απώλειες ενεργού ισχύος. Ο αλγόριθμος μάθησης εφαρμόστηκε για τη βέλτιστη τοπολογία του ΔΔΗΕ εφαρμογής των 33 ζυγών. Τα αποτελέσματα έδειξαν ότι ο προτεινόμενος αλγόριθμος ΕΜ είναι ικανός να παρέχει τη βέλτιστη τοπολογία από ότι άλλοι εξελικτικοί αλγόριθμοι ΤΝ. Επιπλέον, ο ΕΜ παρέχει on-line βέλτιστη τοπολογία του ΔΔΗΕ των 33 ζυγών καθ' όλη τη διάρκεια της λειτουργικής περιόδου.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Venkatesh, B., Ranjan, R.: **Optimal Radial Distribution System Reconfiguration using Fuzzy Adaption of Evolutionary Programming. Int.** J. Electrical Power & Energy Systems, 2003, 25, 75-780.

[2] Baran, M.E., Wu, F.F.: **Network reconfiguration in distribution systems for loss reduction and load balancing**. IEEE Trans. on Power Delivery, 1989, 4, 1401-1407.

[3] Shirmohamaddi, D., Hong, H.W.: **Reconfiguration of electric distribution networks for resistive line losses reduction**. IEEE Trans. on Power Delivery, 1989, 4, 1484-1491.

[4] Peponis, G.P., Papadopoulos, M.P., Hatziargyriou, N.D.: **Distribution networks reconfiguration to minimize resistive line losses. IEEE Trans. on Power Delivery**, 1995, 10, 1338-1342.

[5] Kashem, M.A., Ganapathy, V., Jasmon, G.B., Buhari, M.I.: **A novel method for loss minimization in distribution networks. Proc of Inter. Conf. on Electric Utility Deregulation and Restruct. and Power Tech.**, London, 2000 251-255.

[6] Watkins, C.J.C.H., Dayan, P.: **Q-learning. Machine Learning**, 1992, 8, 279-292.

[7] Kaelbling, L.P., Littman, M.L., Moore, A.W.: **Reinforcement Learning: A Survey**. Journal of Artificial Intelligence Research, 1996, 4, 237-285.

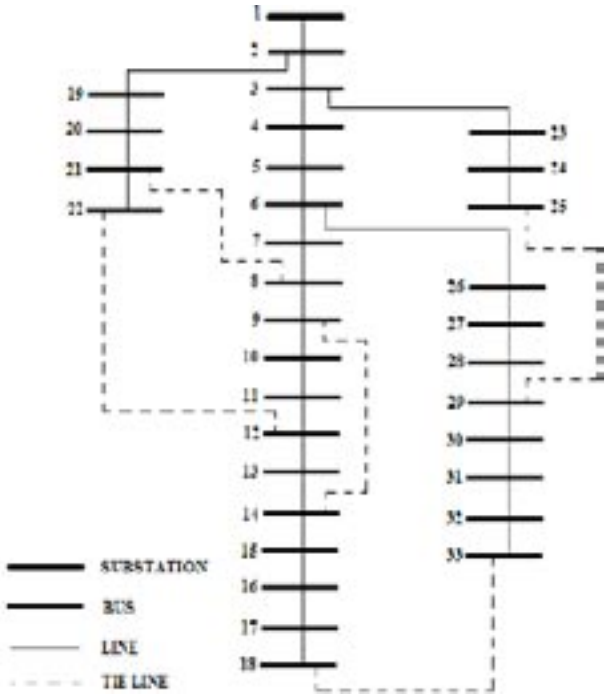
[8] Sutton, R.S., Barto, A.G.: **Reinforcement Learning: An Introduction, Adaptive Computations and Machine Learning**. MIT Press Cambridge MA 1998.

[9] Bertsekas, D.P., Tsitsiklis, J.N.: **Neuro-Dynamic Programming**. Athena Scientific Belmont MA 1996.

[10] Vlachogiannis, J.G., Hatziargyriou, N.D.: **Reinforcement learning (RL) to optimal reconfiguration of radial distribution system (RDS)**, Lecture Notes in Artificial Intelligence, 2004, 3025, 439-446.

ΠΑΡΑΡΤΗΜΑ

Μονογραμμικό διάγραμμα του δικτύου διανομής ηλεκτρικής ενέργειας (ΔΔΗΕ) των 33 ζυγών.



Ιωάννης Γ. Βλαχογιάννης
Δρ Ηλεκτρολόγος Μηχανικός Α.Π.Θ., Υπεύθυνος Εργαστηρίου Βιομηχανικής και Ενεργειακής Πληροφορικής, Σ.Σ. Λιανοκλαδίου, 35100, Λαμία, Email: vlachogiannis@usa.com

Νίκος Δ. Χατζηαργυρίου
Καθηγητής ΕΜΠ, Διευθυντής Εργαστηρίου Ηλεκτρικής Ισχύος, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΕΜΠ, Πολυτεχνειούπολη, Ζωγράφου, Αθήνα, Email: nh@power.ece.ntua.gr

Extended Summary

Optimal Reconfiguration of Radial Distribution System using Machine Reinforcement Learning

J. G. VLACHOGIANNIS
Dr Electrical Engineering AUT

N. D. HATZIARGYRIOU
Professor N.T.U.A

Abstract
This paper presents a Reinforcement Learning (RL) method for optimal reconfiguration of a radial distribution system (RDS). Optimal reconfiguration involves selection of the best set of branches to be opened, one from each loop, such that the resulting RDS has the desired performance. Among the several performance criteria considered for optimal network reconfiguration, an important one is the minimization of real power losses while satisfying voltage limits. The RL method formulates the reconfiguration of RDS as a multistage decision problem. More specifically, the model-free learning algorithm (Q-learning) learns by experience how to adjust a closed-loop control rule, mapping operating states to control actions by means of reward values. Rewards are chosen to express how well control actions cause minimization of power losses. The Q-learning algorithm was applied to the reconfiguration of a 33-bus RDS busbar system. The results are compared with those given by other evolutionary programming methods.

INTRODUCTION

The reconfiguration of a radial distribution system (RDS) aims at its optimal operation, satisfying physical and operating constraints. One of the criteria for optimal operation is the minimization of the real power losses, while simultaneously satisfying operating voltage limits. A number of algorithms based on evolutionary computation techniques have been developed to solve this problem. These methods, however, are inefficient in providing optimal configurations for a whole planning period. In this paper the RDS problem is solved by means of Reinforcement Learning (RL). RL originates from optimal control theory and dynamic programming and aims at approximating solutions to problems of unknown dynamics based on experience. From a theoretical point of view, many breakthroughs have been realized concerning the convergence of the RL approach and their application to nonlinear systems, leading to very efficient algorithms. Also, the rapid increase in computer capacities makes RL methods feasible and attractive in the power system community. Reinforcement Learning (RL) techniques are simple iterative algorithms that learn to act in an optimal way through experience gained by exploring an unknown system. RL assumes that the "world" can be described by a set of states S and an "agent" can choose one action from a set of actions A. The operating range is divided into discrete learning-steps. At each learning-step the agent observes the current state s of the "world" ($s \in S$), and chooses an action $a \in A$ that tends to maximize an expected long-term value function. After taking action (a), the agent is given an immediate reward $r \in \Re$, expressing the effectiveness of the action and observing the resulting state of the "world" $s' \in S$. The particular RL algorithm used in this work is the Q-learning algorithm. In order to apply Q-learning, the reconfiguration problem of RDS is formulated as a multistage decision problem. Optimal control settings are learnt by experience adjusting a closed-loop control rule, which maps operating states to control actions (a set of branches switched off one by one at each loop of RDS). The control settings are based on rewards, expressing how well actions work over the whole planning period. The real power losses function is chosen as reward. Moreover, all voltage limits must be satisfied. Although in this paper the model-free learning algorithm (Q-learning) is applied to optimal reconfiguration of RDS, the algorithm is general and can be applied to a wide variety of optimization problems in planning or operational planning. For the purpose of our analysis a two-class classification is assumed. The "world" of RL states is binary, comprising operating points characterized by satisfaction of their constraints and operating points which violate their constraints. The control vectors that combine branches to be opened one by one at each loop of RDS are the actions, and the Q-learning algorithm is the "agent". The Q-learning algorithm proceeds as follows: An operating point comprising a load and generation pattern including a set of control actions is created randomly. The agent observes the state (s) of the system, as obtained by the load flow solution, and chooses one control action (a) from the control vector. A new load flow is executed. The agent observes the resulting state of the solution (s´) and

provides an immediate reward (r), expressing the reduction of power losses. A new control (switching) action is selected next, leading to a new load flow solution and a new reward. Selection of new control actions is repeated until no more changes in the reward value or in control (switching) action can be achieved. The goal of the agent is to learn the optimal Q-function using the mappings of states to actions such that the long-term reward is maximized. The procedure is repeated for a large number of operating states covering the whole planning period. The agent finds the optimal control settings (a*) using the optimal policy.

The paper is organized in 4 sections. Section 1 describes the Reinforcement Learning approach. In Section 2, the Q-learning algorithm is applied to optimal reconfiguration of RDS. In Section 3, the results obtained by the application of the Q-learning algorithm to the 33-bus RDS are presented. The results are compared with those obtained by the evolutionary programming algorithm, showing the superiority of RL. In addition, the superiority of the Q-learning algorithm in providing optimal reconfiguration over the whole planning period is demonstrated. In Section 4, general conclusions are drawn.

Dr John G. Vlachogiannis
Industrial and Energy Informatics Laboratory (IEI-Lab), R.S. Lianokladiou, 35100, Lamia, Greece, Email: vlachogiannis@usa.com
Professor Nikos D. Hatziargyriou
Power System Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), Polytechnioupoli Zografou, Athens, Greece, Email: nh@power.ece.ntua.gr