

Χρονικά μεταβαλλόμενα πρότυπα

Εισαγωγή

Πρότυπα τα οποία χαρακτηρίζονται από μία καθορισμένη χρονική μεταβολή παρουσιάζονται συχνότατα στις ανθρώπινες δραστηριότητες. Παραδείγματα τέτοιων προτύπων υπάρχουν πολλά όπως η ομιλία, η αντίληψη κινούμενων αντικειμένων, η γραπτή γλώσσα, η εξέλιξη συμπτωμάτων ασθενειών, των φυσικών και εργαστηριακών μετρήσεων σε ασθενείς, η μεταβολή της ονομαστικής αξίας μετοχών κ.ο.κ.

Τα χρονικά μεταβαλλόμενα πρότυπα δεν μπορούν να αντιμετωπιστούν σαν τα απλά πρότυπα που έχουμε ήδη μελετήσει διότι κατά τις διαφορετικές εμφανίσεις τους παρουσιάζουν εκτός από την διασπορά των παραμέτρων τους στον χώρο των μετρήσεων και μία διασπορά (διαφορετική χρονική εξέλιξη) στο πεδίο του χρόνου. Όλα τα παραδείγματα που αναφέραμε χαρακτηρίζονται από σημαντική διαφοροποίηση στην ταχύτητα εξέλιξής τους στο πεδίο του χρόνου, π.χ. κάποιος που βιάζεται να μας αναγγείλει ένα σημαντικό μήνυμα προφέρει τις λέξεις με μεγαλύτερη ταχύτητα από την συνήθη τρόπο ομιλία του, με συνέπεια τα φωνήματα που απαρτίζουν την ομιλία του να έχουν μικρότερη διάρκεια.

Τρεις είναι οι σημαντικότεροι μέθοδοι προσομοίωσης προτύπων που μεταβάλλονται στον χρόνο. Οι μέθοδοι χρονικής ευθυγράμμισης που βασίζονται σε τεχνικές δυναμικού προγραμματισμού, τα στοχαστικά μοντέλα με κυρίαρχα το μοντέλο Markov και το κρυμμένο μοντέλο Markov, και τα ανατροφοδοτούμενα νευρωνικά δίκτυα. Στα πλαίσια του μαθήματος θα δώσουμε μερικά εισαγωγικά στοιχεία από τις μεθόδους που χρησιμοποιούμε όταν ταξινομούμε πρότυπα που χαρακτηρίζονται και

από μεταβολές των προτύπων στο πεδίο του χρόνου.

Αποκρατικές μέθοδοι

Ας ξεκινήσουμε από ένα απλό παράδειγμα. Εστω ότι έχουμε στην διάθεσή μας δύο ακολουθίες παρατηρήσεων δυαδικών αριθμών, την 00111 και την 010101. Οι ακολουθίες αυτές έστω ότι εκφράζουν δύο κατηγορίες γεγονότων την ω_1 και την ω_2 .

Εστω ότι μας δίνεται η ακολουθία γεγονότων 0111. Από ποια κατηγορία γεγονότων έχει παραχθεί η παρατηρούμενη ακολουθία;

Αρχικά παρατηρούμε ότι χρειαζόμαστε μία μέθοδο μέτρησης της απόστασης δύο χρονοσειρών. Αν είχαμε στην διάθεσή μας μία συνάρτηση η οποία να υπολογίζει το μέγεθος $D(00111,0111)$ και την $D(010101,0111)$ τότε η κατηγορία με την μικρότερη απόσταση από το άγνωστο πρότυπο θα ήταν μία βέλτιστη λύση.

Η εύρεση της συνολικής απόστασης του μεγέθους $D(a,b)$ όπου a,b είναι συμβολοσειρές αυθαίρετου μήκους απαιτεί την γνώση του κόστους αντικατάστασης παρεμβολής και και διαγραφής για κάθε ένα από τα σύμβολα.

Εστω λοιπόν ότι είναι γνωστό ότι το κόστος αντικατάστασης του συμβόλου 0 από το 1 και το αντίθετο, της αντικατάστασης του 1 από το 0, είναι 1. Επίσης γνωρίζουμε ότι το κόστος παρεμβολής και διαγραφής συμβόλου είναι 0.5.

Συνεπώς το πρόβλημα που καλούμαστε να λύσουμε είναι η εύρεση του ελάχιστου αριθμού στοιχειωδών μετασχηματισμών με τους οποίους μπορούμε να μεταβούμε από την μία συμβολοσειρά στην δεύτερη. Για να υπολογίσουμε το συνολικό κόστος αυτών των μετασχηματισμών χρησιμοποιούμε τις τεχνικές του δυναμικού προγραμματισμού οι οποίες δίνουν αναδρομικές σχέσεις υπολογισμού της συνολικής απόστασης συμβολοσειρών.

Δυναμικός προγραμματισμός

Εστω ότι θέλουμε να υπολογίσουμε την απόσταση $D(a_1a_2\dots a_x, b_1b_2\dots b_y)$ όπου a_i και b_j είναι σύμβολα, ενώ γνωρίζουμε τις αποστάσεις των $D(a_1a_2\dots a_{x-1}, b_1b_2\dots b_y)$, $D(a_1a_2\dots a_x, b_1b_2\dots b_{y-1})$ και $D(a_1a_2\dots a_{x-1}, b_1b_2\dots b_{y-1})$. Χρησιμοποιώντας την αρχή του ελάχιστου κόστους, η συνολική απόσταση δίνεται από την σχέση:

$$D(a_1a_2\dots a_x, b_1b_2\dots b_y) = \min \begin{cases} D(a_1a_2\dots a_{x-1}, b_1b_2\dots b_y) + d_p(a_x) \\ D(a_1a_2\dots a_{x-1}, b_1b_2\dots b_y) + d(a_x, b_y) \\ D(a_1a_2\dots a_x, b_1b_2\dots b_{y-1}) + d_d(b_y) \end{cases}$$

Από την αναδρομική σχέση βλέπουμε ότι, αν έχουμε στην διάθεσή μας τις αριθμητικές τιμές του κόστους των αντικαταστάσεων των παραμβολών και των

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

διαγραφών συμβόλων μπορούμε να υπολογίσουμε το συνολικό μέγεθος της απόστασης συμβολοσειρών.

Ας δούμε, για το συγκεκριμένο παράδειγμα που μελετάμε, πως βρίσκουμε την συνολική απόσταση. Εστω ότι γνωρίζουμε το κόστος των μετασηματισμών συμβόλων και για χάρη απλότητας των υπολογισμών θέτουμε τις ακόλουθες τιμές:

	$d_p(*),$ παρεμβολή	$d_a(*),$ διαγραφή	$d(0,*),$ αντικατ.	$d(1,*),$ αντικατ.
0	1	1	2	2
1	1	1	2	2

Εστω λοιπόν ότι θέλω να υπολογίσω την συνολική απόσταση $D(00111,0111)$.

Κατασκευάζουμε τον πίνακα που ακολουθεί, τοποθετώντας τα στοιχεία των συμβόλων στην πρώτη γραμμή και πρώτη στήλη αντίστοιχα χωρίς να συμπληρώσω τα στοιχεία των δύο πρώτων κελιών. Στην συνέχεια συμπληρώνω την δεύτερη γραμμή και στήλη με την αντίστοιχη αθροιστική απόσταση σαν το άθροισμα κόστους παρεμβολών και διαγραφών αντίστοιχα για τις συμβολοσειρές αρχίζοντας όπως είναι φυσικό με συνολική απόσταση 0 για το στοιχείο (1,1) του πίνακα. Το αποτέλεσμα των υπολογισμών που κάνω για το συγκεκριμένο παράδειγμα και για κόστος παρεμβολής και διαγραφής ίσο με την μονάδα είναι το ακόλουθο:

		0	0	1	1	1
	0	1	2	3	4	5
0	1					
1	2					
1	3					
1	4					

Στην συνέχεια κατά αυξανόμενο αριθμό στηλών, ή αυξανόμενο αριθμό γραμμών, ή αυξανόμενο αριθμό δευτερευόντων διαγωνίων, χρησιμοποιώντας την αναδρομική σχέση εκτίμησης της συνολικής απόστασης, συμπληρώνω τα υπόλοιπα στοιχεία του πίνακα. Μετά την συμπλήρωση θα έχω το ακόλουθο αποτέλεσμα.

		0	0	1	1	1
	0	1	2	3	4	5
0	1	0	1	2	3	4
1	2	1	2	1	2	3
1	3	2	3	2	1	2
1	4	3	4	3	2	1

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

Η συνολική απόσταση $D(00111,0111)$ είναι ίση με το αριθμό που βρίσκεται στο κάτω δεξιά τετράγωνο, δηλαδή $D(00111,0111) = 1$.

Με την ίδια μέθοδο υπολογίζουμε και την συνολική απόσταση $D(010101,0111)$ κατασκευάζοντας τον αντίστοιχο πίνακα

		0	1	0	1	0	1
	0	1	2	3	4	5	6
0	1	0	1	2	3	4	5
1	2	1	0	1	2	3	4
1	3	2	1	2	3	4	5
1	4	3	2	3	2	3	4

Η συνολική απόσταση είναι $D(010101,0111) = 4$.

Επειδή ισχύει $1 = D(00111,0111) < D(010101,0111) = 4$ το άγνωστο πρότυπο ταξινομείται στα πρότυπα της πρώτης κατηγορίας διότι έχει την μικρότερη αθροιστική απόσταση από το παράδειγμα που διαθέτουμε.

Γενικότερα στην περίπτωση κατά την οποία διαθέτουμε πολλαπλά παραδείγματα ανά κατηγορία η ταξινόμηση πραγματοποιείται με αναζήτηση εκείνης της κατηγορίας που διαθέτει το πρότυπο το οποίο παρουσιάζει την μικρότερη απόσταση.

Παράδειγμα. Έστω ότι διαθέτουμε τα ακόλουθα παραδείγματα δύο κατηγοριών συμβολοσειρών 0 και 1:

$$\Omega_1 = \{0101, 1010\}$$

$$\Omega_2 = \{000, 11111\}$$

Μας δίνεται η συμβολοσειρά 0011, σε ποια κατηγορία προτύπων θα την ταξινομήσετε γνωρίζοντας ότι το κόστος των μειοψηφισμάτων είναι:

	$d_p(*)$	$d_d(*)$	$d(0,*)$	$d(1,*)$
0	2	2	0.5	0.5
1	2	2	0.5	0.5

Εύρεση του βέλτιστου μονοπατιού μετασχηματισμών

Αν θέλουμε να βρούμε τους μετασχηματισμούς με τους οποίους οδηγούμαστε στον υπολογισμό της συνολικής ελάχιστης απόστασης τότε χρησιμοποιούμε τον πίνακα που έχουμε κατασκευάσει. Υπολογίζουμε το βέλτιστο μονοπάτι εφαρμόζοντας την ακόλουθη αναδρομική σχέση:

$$(x^*_{t+1}, y^*_{t+1}) = \arg \min \begin{cases} D(x^*_t - 1, y^*_t) \\ D(x^*_t - 1, y^*_t - 1) \\ D(x^*_t, y^*_t - 1) \end{cases}$$

Όπου $D(x^*_t, y^*_t)$ είναι το περιεχόμενο του κελιού (x^*_t, y^*_t) . Αρχική τιμή θέτουμε το κάτω δεξιό κελί του πίνακα. Όπου (x^*_t, y^*_t) είναι οι βέλτιστες συντεταγμένες που έχουν βρεθεί την χρονική στιγμή t .

Χρησιμοποιώντας τον πίνακα του προηγούμενου παραδείγματος χρωματίζουμε το μονοπάτι που θα προκύψει από την αναδρομική σχέση εκτίμησης του βέλτιστου μονοπατιού μετασχηματισμών

		0	1	0	1	0	1
	0	1	2	3	4	5	6
0	1	0	1	2	3	4	5
1	2	1	0	1	2	3	4
1	3	2	1	2	3	4	5
1	4	3	2	3	2	3	4

Από τα αποτελέσματα βλέπουμε ότι το βέλτιστο μονοπάτι απαρτίζεται από δύο παρεμβολές και τέσσερις αντικαταστάσεις.

Παράδειγμα. Για το πρότυπο του προηγούμενου παραδείγματος το οποίο έχει την μικρότερη απόσταση βρέται το αντίστοιχο βέλτιστο μονοπάτι μετασχηματισμών.

Εκτίμηση κόστους μετασχηματισμών αντικατάστασης, παρεμβολής, διαγραφής συμβόλων

Για τον υπολογισμό της συνολικής απόστασης συμβολοσειρών και κατά συνέπεια και για τον υπολογισμό του βέλτιστου μονοπατιού μετασχηματισμών χρησιμοποιήσαμε τα

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

μεγέθη κόστους τα οποία θεωρήσαμε ότι είναι γνωστά.

Στις περισσότερες των εφαρμογών τα δεδομένα αυτά είναι γνωστά ή υπολογίζονται ανεξάρτητα (από την φύση του προβλήματος). Χαρακτηριστικό παράδειγμα αποτελούν τα συστήματα αυτόματης διόρθωσης σφαλμάτων. Σε αυτή την περίπτωση κατηγορίες προτύπων είναι οι λέξεις του διορθωτή κειμένου (έχουμε ένα παράδειγμα ανά κατηγορία προτύπων, την ορθογραφική γραφή της λέξης). Το πρόβλημα που πρέπει να λύσουμε είναι να διορθωθεί η λέξη που δίνεται (π.χ. η λέξη "φοιτιτίς" πρέπει να ταξινομηθεί στην κατηγορία που περιέχει την λέξη "φοιτητής"). Το κόστος των μετασχηματισμών αντικατάστασης, παραμβολής και διαγραφής συσχετίζεται με την συχνότητα των λαθών και όπως είναι φυσικό τα κόστη αυτά να εξαρτώνται από τον τρόπο παραγωγής του κειμένου. Αν το κείμενο έχει παραχθεί με πληκτρολόγηση, το κόστος αντικατάστασης $d(1,i)$ είναι μεγάλο διότι πολύ δύσκολα ο άνθρωπος που πληκτρολόγησε το κείμενο μπορεί να μπερδέψει το πλήκτρο 1 με το πλήκτρο i διότι αυτά είναι απομακρυσμένα στο πληκτρολόγιο. Στην περίπτωση όμως κατά την οποία το κείμενο αποτελεί την έξοδο συστήματος οπτικής αναγνώρισης χαρακτήρων, το κόστος αντικατάστασης γίνεται πολύ μικρό διότι οι χαρακτήρες 1 και i μοιάζουν πολύ και είναι πιθανόν το σύστημα οπτικής αναγνώρισης χαρακτήρων να κάνει λανθασμένη ταξινόμηση.

Αν δεν γνωρίζουμε τα μεγέθη κόστους αλλά διαθέτουμε μόνο ένα πλήθος από παραδείγματα, μπορούμε να εκτιμήσουμε την αριθμητική τιμή κόστους των μετασχηματισμών ελαχιστοποιώντας την αθροιστική τους απόσταση. Δηλαδή έστω ότι διαθέτουμε N παραδείγματα συμβολοσειρών:

$$\Omega = \{\pi_1, \pi_2, \dots, \pi_N\}$$

Εφαρμόζοντας ακόλουθο τον επαναληπτικό αλγόριθμο μπορούμε να εκτιμήσουμε την αριθμητική τιμή του κόστους των μετασχηματισμών, ελαχιστοποιώντας το ακόλουθο άθροισμα

$$D = \sum_{i \neq j} D(\pi_i, \pi_j)$$

Αλγόριθμος Εκτίμησης κόστους μετασχηματισμών

Αρχικοποίηση

Τοποθετώ αρχικές τιμές στο κόστος των μετασχηματισμών. Οι τιμές που επιλέγω μπορεί να είναι τυχαίες αλλά για την ταχύτερη σύγκλιση του επαναληπτικού τμήματος του αλγόριθμου τοποθετώ σαν κόστος αντικατάστασης την τιμή 2 και σαν κόστος παραμβολής και διαγραφής την τιμή 1.

Επαναληπτικό τμήμα του αλγόριθμου. Για κάθε ένα από τα $N(N-1)/2$ ζεύγη συμβολοσειρών από τα παραδείγματα που διαθέτω υπολογίζω το βέλτιστο μονοπάτι μετασχηματισμών.

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

Επαναπροσδιορισμός του κόστους των μετασχηματισμών. Από το σύνολο των βέλτιστων μετασχηματισμών υπολογίζω την συχνότητα εμφάνισης κάθε ενός από αυτούς. Από την συχνότητα εμφάνισης υπολογίζω την αντίστοιχη πιθανότητα.

Για κάθε μετασχηματισμό που δεν εμφανίζεται έστω και μία φορά στα βέλτιστα μονοπάτια μετασχηματισμών, υποθέτω συχνότητα εμφάνισης ένα.

Η νέα εκτίμηση του κόστους των μετασχηματισμών προκύπτει από την αντιστροφή των πιθανοτήτων εμφάνισης:

$$d(x, y) = (p(x, y))^{-1} = \frac{Ns}{N(x, y)}$$

όπου Ns είναι το πλήθος των μετασχηματισμών.

Έλεγχος σύγκλισης. Αν οι νέοι υπολογισμοί του κόστους των μετασχηματισμών έχει διαφοροποιηθεί σημαντικά από την προηγούμενη εκτίμησή του τότε οι νέες τιμές του κόστους αντικαθιστούν τις παλιές εκτιμήσεις και το επαναληπτικό τμήμα του αλγόριθμου επαναλαμβάνεται, διαφορετικά ο αλγόριθμος τερματίζει.

Παράδειγμα. Έστω ότι διαθέτουμε τα ακόλουθα παραδείγματα

$$\Omega_1 = \{0101, 1010, 000\}$$

Υλοποιήσε τον αλγόριθμο εκτίμησης του κόστους των μετασχηματισμών μόνο για την πρώτη επανάληψη του αλγόριθμου, θέτοντας σαν αρχική εκτίμηση τις τιμές που δίνονται στον ακόλουθο πίνακα

	$d_b(^*)$	$d_d(^*)$	$d(0,^*)$	$d(1,^*)$
0	2	2	0.5	0.5
1	2	2	0.5	0.5

Στοχαστικά συστήματα

Τα στοχαστικά συστήματα διακρίνονται για την απλότητά τους και την περιγραφική τους ακρίβεια. Σε αντίθεση με τα νευρωνικά δίκτυα στα οποία δεν μπορούμε να προσδώσουμε ένα πλήρες φυσικό νόημα στις σταθερές παραμέτρους του δικτύου, στην περίπτωση των στοχαστικών συστημάτων κάθε παράμετρος του μοντέλου αντιστοιχεί σε μία πιθανότητα που περιγράφει στοχαστική συμπεριφορά σε μικροσκοπική κλίμακα.

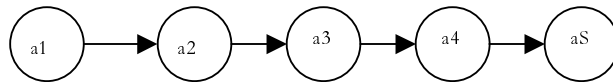
Η πλέον απλή στοχαστική περιγραφή χρονικά μεταβαλλόμενων φαινομένων επιτυγχάνεται με τις αλυσίδες Markov.

Το μοντέλο του Markov

Έστω λοιπόν ότι έχουμε μία πηγή η οποία σε τακτά χρονικά διαστήματα παράγει ένα σύμβολο το οποίο ανήκει σε ένα πεπερασμένο σύνολο N συμβόλων, έστω το:

$$\Omega = \{ a_1, a_2, \dots, a_N \}$$

Για να μπορούμε να ισχυριστούμε ότι γνωρίζουμε την στοχαστική συμπεριφορά της πηγής θα πρέπει να είμαστε σε θέση να υπολογίσουμε την πιθανότητα παραγωγής οποιασδήποτε ακολουθίας συμβόλων.



Στην πλέον απλή περίπτωση κατά την οποία κρατάμε την αριθμητική τιμή των πιθανοτήτων σε ένα πίνακα τιμών, τότε η μνήμη που απαιτείται για να περιγράψουμε πλήρως την πηγή για μήκος ακολουθίας M συμβόλων, είναι:

$$\text{Αριθμός παραμέτρων} = \sum_{i=1}^M N^i$$

Αν υποθέσουμε λοιπόν ότι θέλουμε να περιγράψουμε πλήρως την στοχαστική συμπεριφορά μίας πηγής η οποία παράγει τέσσερα διαφορετικά σύμβολα με μέγιστο μήκος 15, θα χρειαστούμε μνήμη η οποία θα πρέπει να αποθηκεύει 357.913.941 πραγματικούς αριθμούς μεγάλης ακρίβειας αφού οι αντίστοιχες πιθανότητες, θα είναι μικροί αριθμοί.

Επίσης χρειαζόμαστε έναν τεράστιο αριθμό παραδειγμάτων έτσι ώστε να εκτιμήσουμε με ικανοποιητική ακρίβεια τις πιθανότητες του μοντέλου. Τόσο μεγάλος αριθμός παραδειγμάτων δεν μπορεί να βρεθεί εύκολα στις πρακτικές εφαρμογές.

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

Από το απλό αυτό παράδειγμα φαίνεται ότι χρειαζόμαστε μία καλύτερη προσέγγιση για την επίλυση του προβλήματος έτσι ώστε να ελαττώσουμε τον αριθμό των παραμέτρων που χρειάζεται η προσομοίωση που επιχειρούμε.

Ας υποθέσουμε λοιπόν ότι το σύμβολο το οποίο παράγεται την χρονική στιγμή t εξαρτάται μονάχα από τα σύμβολα που παρήγαγε η πηγή τις K προηγούμενες χρονικές στιγμές. Τα σύμβολα που παρήγαγε η πηγή σε παρελθόντα χρόνο μεγαλύτερο του K είναι στοχαστικά ασυσχέτιστο με το σύμβολο a_t .

Λαμβάνοντας υπόψη την παραπάνω υπόθεση η πιθανότητα παραγωγής οποιασδήποτε ακολουθίας συμβόλων με μήκος μεγαλύτερο του K μπορεί να απλοποιηθεί ως εξής.

Η αρχική πιθανότητα γράφεται:

$$p(a_1 a_2 \dots a_S) = p(a_1 a_2 \dots a_{S-1}) p(a_S | a_1 a_2 \dots a_{S-1})$$

Το σύμβολο a_S εξαρτάται μονάχα από την ακολουθία των K προηγούμενων συμβόλων, συνεπώς η δεύτερη πιθανότητα απλοποιείται σε:

$$p(a_S | a_1 a_2 \dots a_{S-1}) = p(a_S | a_{S-K+1} \dots a_{S-1})$$

Αν αντικαταστήσουμε στην εξίσωση τον απλοποιημένο όρο και επαναλάβουμε την ίδια διαδικασία για την πρώτη πιθανότητα του γινομένου, τότε η πιθανότητα $p(a_1 a_2 \dots a_S)$ απλοποιείται σε:

$$p(a_1 a_2 a_3 \dots a_S) = p(a_1 a_2 a_3 \dots a_K) \prod_{i=K+1}^S p(a_i | a_{i-K} a_{i-K+1} \dots a_{i-1})$$

Από την παραπάνω εξίσωση γίνεται κατανοητό ότι μπορούμε να περιγράψουμε πλήρως οποιαδήποτε συμβολοσειρά της πηγής χρησιμοποιώντας τον ακόλουθο αριθμό παραμέτρων:

$$\text{Αριθμός Παραμέτρων} = \sum_{i=1}^K N^i + N^{K+1}$$

για N σύμβολα πηγής και M μήκος ακολουθίας συμβόλων:

Η βελτίωση που επιφέραμε στο μοντέλο επικεντρώνεται στο γεγονός ότι η μνήμη που απαιτείται για να περιγράψουμε οποιαδήποτε συμβολοσειρά δεν εξαρτάται πλέον από το

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

μήκος της αλλά μονάχα από το μήκος της στοχαστικής εξάρτησης K και τον αριθμό των συμβόλων N . Το σημαντικό αυτό επίτευγμα μας οδηγεί στο συμπέρασμα ότι μπορούμε να υπολογίσουμε την πιθανότητα παραγωγής οποιασδήποτε συμβολοσειράς, ανεξαρτήτου μήκους. Η προσομοίωση που πετύχαμε ονομάζεται μοντέλο Markov K -τάξης.

Έτσι στο αριθμητικό παράδειγμα που θέσαμε, αν υποθέσουμε ότι η τάξη του μοντέλου Markov είναι πέντε ($K=5$) τότε με την γνώση 4437 πιθανοτήτων μπορούμε να υπολογίσουμε την πιθανότητα παραγωγής οποιασδήποτε συμβολοσειράς.

Η εκτίμηση των πιθανοτήτων κάθε μοντέλου από παραδείγματα μπορεί να πραγματοποιηθεί εύκολα υπολογίζοντας τις αντίστοιχες στατιστικές συχνότητες εμφάνισης ακολουθιών συμβόλων από παραδείγματα.

Παράδειγμα. Έστω ότι διαθέτουμε τα εξής παραδείγματα από δύο διαφορετικές πηγές G_1 και G_2 δυαδικών συμβόλων:

$$\Omega_1 = \{ 10101, 010101, 001101, 01010 \}$$

$$\Omega_2 = \{ 00001, 1000000, 0000, 00010, 111111 \}$$

1. Υποθέτοντας ότι οι ακολουθίες συμβόλων προέρχονται από πηγές Markov πρώτης τάξης, υπολογίστε τις πιθανότητες του πλήρους μοντέλου, και ταξινομήστε την ακολουθία 00100 υπολογίζοντας τις δεσμευμένες πιθανότητες $p(00100|G_1)$ και $p(00100|G_2)$.
2. Αν ενσωματώσετε την ακολουθία που ταξινομήσατε στα παραδείγματα της πηγής η οποία την παράγει με την μεγαλύτερη πιθανότητα και εκπαιδεύσετε ξανά το μοντέλο η θα συμβεί στην πιθανότητα παραγωγής της συμβολοσειράς; Δώστε μία εξήγηση και επιβεβαιώστε την πρόβλεψή σας.

Το κρυμμένο μοντέλο Markov

Πολλές φορές τα προβλήματα που συναντάμε δεν είναι τόσο απλά. Θόρυβος ο οποίος παρεμβάλλεται στα μετρητικά μας όργανα αλλοιώνει τις μετρήσεις μας με αποτέλεσμα τα δεδομένα μας να μην αντιστοιχούν στα πραγματικά σύμβολα των πηγών.

Για να γίνει κατανοητό το πρόβλημα θα δώσουμε ένα παράδειγμα που αποτελεί μία πιο περίπλοκη κατάσταση του προηγούμενου παραδείγματος. Έστω ότι οι πηγές συμβολοσειρών που περιγράψαμε βρίσκονται εγκατεστημένες σε δορυφόρο ο οποίος στέλνει μηνύματα σε δέκτη ο οποίος είναι τοποθετημένος στην Γη. Υποθέτουμε ότι οι ακολουθίες συμβόλων που παράγονται στον δορυφόρο προέρχονται από πηγές Markov πρώτης τάξης. Ο δορυφόρος εκπέμπει τα δυαδικά σύμβολα αλλά ατμοσφαιρικός θόρυβος τα παραμορφώνει. Υποθέτοντας ότι η στοχαστική διαδικασία

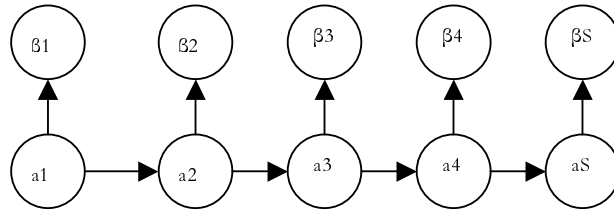
ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

παραμόρφωσης είναι *χρονικά αμετάβλητη* και εξαρτάται μονάχα από το σύμβολο που εκπέμπει ο δορυφόρος, μπορούμε να βρούμε ποιά πηγή εξέπεμψε σήμα το οποίο στον δέκτη της Γης αναγνωρίστηκε σαν $\beta_1, \beta_2, \dots, \beta_S$;

Το ενδιαφέρον αυτό πρόβλημα, το οποίο είναι πληρέστερο για πρακτικές εφαρμογές στις οποίες υπάρχει θόρυβος που αλλοιώνει τις μετρήσεις μας, έχει λύση. Η λύση θα είναι πλήρης αν με την γνώση των πιθανοτήτων του μοντέλου μπορούμε να υπολογίσουμε την πιθανότητα $p(\beta_1, \beta_2, \dots, \beta_S)$.

Ας ξεφύγουμε λίγο από το παράδειγμα και ας προσπαθήσουμε να αντιμετωπίσουμε το πρόβλημα στην γενικοτήτά του.

Έστω λοιπόν ότι έχουμε μία πηγή Markov K-τάξης. Τα σύμβολα της πηγής τα οποία δεν είναι άμεσα παρατηρήσιμα και ονομάζονται κρυφές καταστάσεις του μοντέλου, μετασχηματίζονται σε μία διαφορετική ακολουθία συμβόλων η οποία είναι άμεσα παρατηρήσιμη και γιαυτό τον λόγο θα ονομάζεται ακολουθία παρατηρήσεων.



Σχήμα 4.1 Σχηματική παράσταση του κρυμμένου μοντέλου Markov

Αν λοιπόν $\Omega_a = \{ a_1, a_2, \dots, a_N \}$ είναι τα N σύμβολα της πηγής, και $\Omega_\beta = \{ \beta_1, \beta_2, \dots, \beta_M \}$ είναι τα M σύμβολα παρατήρησης τότε κάνοντας τις ακόλουθες υποθέσεις:

- Η ακολουθία συμβόλων της πηγής ή των κρυφών καταστάσεων μπορεί να προσομοιωθεί με μία πηγή Markov K-τάξης,
- σε κάθε χρονική στιγμή το σύμβολο της κρυφής κατάστασης παράγει ένα αντίστοιχο σύμβολο παρατήρησης,
- το σύμβολο παρατήρησης εξαρτάται στοχαστικά μόνο από το αντίστοιχο σύμβολο παρατήρησης στο κρυφό επίπεδο της ίδιας χρονικής στιγμής,
- τα στοχαστικά χαρακτηριστικά του μοντέλου προσομοίωσης παραμένουν χρονικά αμετάβλητα,

ας δούμε αν μπορούμε να υπολογίσουμε την πιθανότητα παρατήρησης μίας ακολουθίας συμβόλων, και τι είδους πιθανότητες χρειάζονται σε έναν τέτοιο υπολογισμό.

Η πιθανότητα παρατήρησης μίας ακολουθίας καταστάσεων είναι:

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

$$p(\beta_1 \beta_2 \dots \beta_S) = \sum_{a_1 a_2 \dots a_S}^R p(\beta_1 \beta_2 \dots \beta_S | a_1 a_2 \dots a_S) p(a_1 a_2 \dots a_S)$$

όπου R είναι ο αριθμός όλων των πιθανών ακολουθιών S συμβόλων στο κρυφό επίπεδο.

Από τις υποθέσεις που έχουμε κάνει για την στοχαστική ανεξαρτησία των συμβόλων που παράγονται στο κρυφό επίπεδο (ακολουθία Markov) η σχέση απλοποιείται σε:

$$p(\beta_1 \beta_2 \dots \beta_S) = \sum_{a_1 a_2 \dots a_S}^R p(\beta_1 \beta_2 \dots \beta_S | a_1 a_2 \dots a_S) p(a_1) \prod_{i=2}^S p(a_i | a_{i-1})$$

Γνωρίζοντας επίσης ότι η παρατήρηση που λαμβάνουμε την χρονική στιγμή t εξαρτάται μόνο από την κρυφή κατάσταση για την ίδια χρονική στιγμή, η δεσμευμένη πιθανότητα απλοποιείται ως εξής:

$$p(\beta_1 \beta_2 \dots \beta_S) = \sum_{a_1 a_2 \dots a_S}^R \prod_{k=1}^S p(\beta_k | a_k) p(a_1) \prod_{i=2}^S p(a_i | a_{i-1})$$

Από την τελευταία σχέση βλέπουμε ότι μπορούμε να υπολογίσουμε την πιθανότητα εμφάνισης οποιασδήποτε συμβολοσειράς $\beta_1 \beta_2 \dots \beta_S$ γνωρίζοντας ένα πεπερασμένο σύνολο πιθανοτήτων. Θα λέμε ότι έχει οριστεί πλήρως η στοχαστική συμπεριφορά μίας διεργασίας που προσομοιώνεται με ένα κρυμμένο μοντέλο Markov όταν είναι γνωστά τα ακόλουθα μεγέθη:

$$\lambda = (\pi, A, B)$$

όπου, $\pi = p(a_i)$, $i=1, N$, είναι η πιθανότητα εκκίνησης της διεργασίας με πρώτο κρυφό σύμβολο το a_i ,

$A = [p(a_i | a_j)]$, $i, j=1, N$, είναι η πιθανότητα να συμβεί το κρυφό γεγονός a_i , ενώ την προηγούμενη χρονική στιγμή έχει συμβεί το γεγονός a_j , και

$B = [p(\beta_i | a_j)]$, $i=1, M$, $j=1, N$, είναι η πιθανότητα να συμβεί η παρατήρηση β_i , ενώ την ίδια χρονική στιγμή έχει συμβεί το κρυφό γεγονός a_j .

Η σημαντική συνεισφορά του κρυμμένου μοντέλου Markov βρίσκεται στο γεγονός ότι με πεπερασμένο αριθμό παραμέτρων μπορούμε να περιγράψουμε χρονικά μεταβαλλόμενες διεργασίες οι οποίες δεν είναι άμεσα παρατηρήσιμες. Αντίθετα από

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

το απλό μοντέλο Markov στο οποίο πρέπει οι παρατηρήσεις που λαμβάνουμε να είναι χωρίς σφάλματα, η διπλή στοχαστική διεργασία του κρυμμένου μοντέλου Markov μπορεί να προσομοιώσει χρονικά μεταβαλλόμενα γεγονότα τα οποία δεν είναι μετρήσιμα με ακρίβεια στον δέκτη. Θόρυβος ή στοχαστικός μετασχηματισμός των γεγονότων, είναι μερικές περιπτώσεις κατά τις οποίες το αρχικό σύνολο παρατηρήσεων παραμορφώνεται από στοχαστικό κανάλι. Οι συνθήκες προσομοίωσης που εισάγει το κρυμμένο μοντέλο Markov ταιριάζει σε πολλά χρονικά μεταβαλλόμενα πρότυπα με συνέπεια, ιδιαίτερα τα τελευταία χρόνια, να πολλαπλασιάζονται οι εφαρμογές του μοντέλου σε συστήματα ταξινόμησης προτύπων, όπως η ομιλία, η οπτική αναγνώριση αντικειμένων, η ιατρική διάγνωση, η ανίχνευση κινούμενων αντικειμένων κ.ο.κ.

Υπολογισμός της πιθανότητας παρατήρησης

Ο αριθμός των πράξεων που απαιτούνται για τον υπολογισμό της πιθανότητας μίας ακολουθίας παρατηρήσεων είναι σχετικά μεγάλος. Με την βοήθεια τεχνικών δυναμικού προγραμματισμού μπορούμε να ελαττώσουμε τις αριθμητικές πράξεις δραστικά, εισάγοντας την αθροιστική πιθανότητα $a_i(i)$, η οποία ορίζεται να είναι η πιθανότητα να λάβουμε τις παρατηρήσεις $\beta_1, \beta_2, \dots, \beta_i$ και την χρονική στιγμή t στο κρυφό επίπεδο να έχει συμβεί η κατάσταση a_i :

$$a_i(i) = p(\beta_1, \beta_2, \dots, \beta_i | a_i)$$

Η δραστική μείωση των υπολογισμών οφείλεται στην ιδιότητα της πιθανότητας $a_i(i)$ να υπολογίζεται με την βοήθεια της αναδρομικής σχέσης:

$$a_i(i) = p(\beta_i | a_i) \sum_{j=1}^N a_{i-1}(j) p(a_i | a_j)$$

Αναλυτικά ο υπολογισμός της πιθανότητας παρατήρησης μίας ακολουθίας γεγονότων $\beta_1, \beta_2, \dots, \beta_N$, δίνεται από τον αλγόριθμο (forward procedure):

Αλγόριθμος

Αρχικοποίηση

$$a_0(j) = \pi_j = p(a_j), j = 1, N$$

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

Επαναληπτικό τμήμα του αλγόριθμου. Υπολογίζεται για $t=1,2,\dots,S$ και $i=1,N$, οι ακόλουθες πιθανότητες:

$$a_t(i) = p(\beta_t | a_t) \sum_{j=1}^N a_{t-1}(j) p(a_t | a_j)$$

Τελικός υπολογισμός. Η πιθανότητα παρατήρησης της ακολουθίας $\beta_1, \beta_2, \dots, \beta_S$, δίνεται από την σχέση:

$$p(\beta_1 \beta_2 \dots \beta_S) = \sum_{j=1}^N a_S(j)$$

Αποδεικνύεται ότι οι υπολογισμοί που απαιτούνται είναι της τάξης του $N^2 S$ ενώ αν χρησιμοποιηθεί η σχέση που προκύπτει από τον αλγόριθμο Viterbi η τάξη μεγέθους των αντίστοιχων υπολογισμών ανέρχεται σε $2 S N^S$.

Υπολογισμός των κρυφών καταστάσεων

Ένα πολύ εύλογο ερώτημα, το οποίο έχει πολλές πρακτικές εφαρμογές, αναφέρεται και στην δυνατότητα να υπολογίσουμε τι πραγματικά συμβαίνει στο κρυφό επίπεδο δοσμένης της ακολουθίας παρατηρήσεων και των παραμέτρων του μοντέλου. Μία διαφορετική κατηγορία προβλημάτων ταξινόμησης προτύπων μπορεί να λυθεί αναγνωρίζοντας την ακολουθία των κρυφών καταστάσεων. Η λύση έχει δοθεί με τον αλγόριθμο Viterbi.

Περιγράφοντας το ερώτημα που τίθεται, ουσιαστικά ζητάμε να λύσουμε την ακόλουθη εξίσωση:

$$(a_{1b} a_{2b} \dots a_{Sb}) = \arg \max_{a_1 a_2 \dots a_S} p(a_1 a_2 \dots a_S | \beta_1 \beta_2 \dots \beta_S)$$

η οποία ισοδυναμεί με το ακόλουθο κριτήριο:

$$(a_{1b} a_{2b} \dots a_{Sb}) = \operatorname{argmax}_{a_1 a_2 \dots a_S} \frac{p(a_1 a_2 \dots a_S | \beta_1 \beta_2 \dots \beta_S)}{p(\beta_1 \beta_2 \dots \beta_S)} = \operatorname{argmax}_{a_1 a_2 \dots a_S} p(a_1 a_2 \dots a_S, \beta_1 \beta_2 \dots \beta_S)$$

Ακολουθώντας την ίδια τεχνική που ακολουθήσαμε για τον υπολογισμό της πιθανότητας των παρατηρήσεων, ορίζουμε την ακόλουθη ποσότητα:

$$\delta_t(i) = \max_{a_1 a_2 \dots a_{t-1}} p(a_1 a_2 \dots a_t = a_t, \beta_1 \beta_2 \dots \beta_S)$$

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

Εύκολα αποδεικνύεται ότι ισχύει η ακόλουθη αναδρομική σχέση:

$$\delta_{i+1}(i) = p(\beta_{i+1} | a_i) \max_j \delta_i(j) p(a_i | a_j)$$

Με την βοήθεια της αναδρομικής εξίσωσης, η συνολική πιθανότητα υπολογίζεται από τον αλγόριθμο Viterbi:

Αλγόριθμος

Αρχικοποίηση.

$$\delta_0(j) = \pi_j = p(a_j), j = 1, N$$

Επαναληπτικό τμήμα του αλγόριθμου. Υπολογίζονται για $t=1,2,\dots,S$ και $i=1,N$, οι ακόλουθες ποσότητες:

$$\delta_t(i) = p(\beta_t | a_i) \max_{1 \leq j \leq N} \delta_{t-1}(j) p(a_i | a_j)$$

$$\psi_t(i) = \arg \max_{1 \leq j \leq N} \delta_{t-1}(j) p(a_i | a_j)$$

Αρχικός υπολογισμός της βέλτιστης κρυφής κατάστασης σε χρόνο S. Η κρυφή κατάσταση $a_\xi(S)$ που μεγιστοποιεί το κριτήριο είναι

$$\xi = \arg \max_i \delta_S(i)$$

Υπολογισμός των κρυφών καταστάσεων. Ο υπολογισμός των κρυφών καταστάσεων πραγματοποιείται από υψηλότερες προς χαμηλότερες αριθμητικές τιμές του χρόνου. Συγκεκριμένα ακολουθώντας την αναδρομική σχέση:

$$\xi_t = \psi_{t+1} \xi_{t+1}$$

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

για τιμές $t=S-1, S-2, \dots, 1$ υπολογίζω την πλέον πιθανή ακολουθία κρυφών καταστάσεων δεδομένης της ακολουθίας παρατηρήσεων και των παραμέτρων του μοντέλου σαν:

$$(a_1, a_2, \dots, a_S) = (a_{\xi_1}, a_{\xi_2}, a_{\xi_3}, \dots, a_{\xi_S})$$

Η εκπαίδευση του κρυμμένου μοντέλου Markov

Είναι γνωστό ότι οποιαδήποτε μέθοδος ταξινόμησης προτύπων μπορεί να τύχει αποδοχής και ευρείας χρήσης σε εφαρμογές αν υπάρχει αλγόριθμος ο οποίος να προσδιορίζει τις παραμέτρους του μοντέλου από ένα σύνολο παραδειγμάτων.

Ο υπολογισμός των παραμέτρων του κρυμμένου μοντέλου Markov από παρατηρήσεις διαφοροποιείται ανάλογα με το είδος της πληροφορίας που είναι διαθέσιμη. Στις επόμενες παραγράφους παρουσιάζονται δύο διαφορετικοί αλγόριθμοι εκπαίδευσης οι οποίοι εφαρμόζονται όταν:

1. είναι διαθέσιμος αριθμός παραδειγμάτων για τα οποία γνωρίζουμε και την ακολουθία κρυφών καταστάσεων και την ακολουθία παρατηρήσεων.
2. είναι διαθέσιμος αριθμός παραδειγμάτων για τα οποία γνωρίζουμε μόνο την ακολουθία παρατηρήσεων.

Είναι προφανές ότι στην πρώτη περίπτωση μπορούμε να εκτιμήσουμε τις παραμέτρους του μοντέλου αν ακολουθήσουμε την μέθοδο στατιστικής προσέγγισης των πιθανοτήτων. Αρχικά υπολογίζουμε τις συχνότητες εμφάνισης που αντιστοιχούν στις πιθανότητες του μοντέλου και στην συνέχεια υπολογίζουμε την στατιστική τους εκτίμηση. Γνωρίζουμε ότι αυξανόμενου του αριθμού παραδειγμάτων η στατιστική εκτίμηση των πιθανοτήτων του μοντέλου προσεγγίζει ασυμπτωτικά τις πραγματικές πιθανότητες.

Παράδειγμα. Επίγειος δορυφορικός σταθμός εδάφους λαμβάνει σήματα από δύο δορυφόρους. Πριν από την αποστολή κάθε μηνύματος ο δορυφόρος στέλνει μία κωδική ακολουθία δυαδικών αριθμών δηλώνοντας την ταυτότητά του. Ο πρώτος στέλνει την ακολουθία 000111 και ο δεύτερος 010101.

Η αιμόσφαιρα παραμορφώνει το σήμα δίνοντας στον επίγειο σταθμό για τον πρώτο δορυφόρο τις ακολουθίες παρατηρήσεων {001111, 100011, 010011} ενώ για τον δεύτερο δορυφόρο καταγράφηκαν τα σήματα {010101, 011101, 110101}.

1. Μπορείτε να προσομοιώσετε την διαδικασία εκπομπής και λήψης σήματος με ένα κρυμμένο μοντέλο Markov;

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

2. Υπολογίστε τις παραμέτρους των κρυφών μοντέλων Markov για το σήμα ταυτότητας που εκπέμπει κάθε δορυφόρος από τις παρατηρήσεις του επίγειου σταθμού.
3. Ποιός δορυφόρος εκπέμπει αν λάβουμε σήμα ταυτότητας το 000000 ;

Η επίλυση του προβλήματος της εκπαίδευσης είναι δυσκολότερη όταν δεν γνωρίζουμε την ακολουθία κρυφών καταστάσεων. Σε αυτή την περίπτωση υπάρχουν διαθέσιμοι δύο αλγόριθμοι. Ο πρώτος ονομάζεται Baum-Welch και βασίζεται στην επιλογή εκείνων των παραμέτρων που μεγιστοποιούν την πιθανότητα

$$\lambda_0 = \max_{\lambda} \prod_{k=1}^I p(O_k | \lambda)$$

όπου O_1, O_2, \dots, O_I είναι τα I παραδείγματα της εκπαίδευσης που αντιστοιχούν σε I ακολουθίες παρατήρησης.

Ο δεύτερος αλγόριθμος ονομάζεται segmental k-means και υπολογίζει τις παραμέτρους του κρυμμένου μοντέλου Markov λύνει με επαναληπτικό τρόπο την εξίσωση:

$$\lambda_0^{(t)} = \arg \max_{\lambda} \left(\prod_{k=1}^I \max_{a_k} p(O_k, a_k | \lambda) \right)$$

Πειραματικές συγκριτικές μελέτες έχουν δείξει ότι και οι δύο λύσεις δίνουν πολύ καλή ακρίβεια στην εκπαίδευση. Ο segmental k-means αλγόριθμος έχει απλούστερη αλγοριθμική υλοποίηση, παρουσιάζει λιγότερα προβλήματα κατά τους υπολογισμούς και είναι ταχύτερος στην πλειονότητα των πρακτικών υλοποιήσεων. Η κύρια αιτία των πλεονεκτημάτων αυτών είναι η θεμελιώδης αρχή του αλγόριθμου που βασίζεται σε μία διπλή επαναληπτική διαδικασία βελτιστοποίησης. Είναι ιδιαίτερα δημοφιλής στις πρακτικές εφαρμογές.

Στα πλαίσια του μαθήματος θα περιγράψουμε μόνο τον αλγόριθμο segmental k-means:

Αλγόριθμος.

Αρχικοποίηση. Επιλέγω με τυχαίο τρόπο τις παραμέτρους του μοντέλου. $t \rightarrow 0$.

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

$$\lambda^{(0)} = (\pi^{(0)}, A^{(0)}, B^{(0)})$$

Τα επόμενα δύο βήματα του αλγόριθμου επαναλαμβάνονται μέχρι να ικανοποιηθεί η συνθήκη του κριτηρίου σύγκλισης.

Εύρεση της βέλτιστης ακολουθίας παρατηρήσεων. Για κάθε μία από τις ακολουθίες παρατηρήσεων βρίσκω την αντίστοιχη ακολουθία κρυφών καταστάσεων χρησιμοποιώντας τις παραμέτρους του μοντέλου $\lambda^{(t)}$ και τον αλγόριθμο Viterbi.

Στατιστική εκτίμηση παραμέτρων. Γνωρίζοντας την κρυφή ακολουθία καταστάσεων και την αντίστοιχη ακολουθία παρατηρήσεων για κάθε παράδειγμα προχωρούμε στην στατιστική εκτίμηση των πιθανοτήτων του μοντέλου.

Έλεγχος σύγκλισης. Υπάρχουν πολλά κριτήρια με τα οποία μπορούμε να μετρήσουμε τον βαθμό ακρίβειας υπολογισμού των παραμέτρων του μοντέλου. Μερικά από αυτά είναι τα ακόλουθα:

Υπολογίζουμε την πιθανότητα να παρατηρήσουμε όλα τα παραδείγματα εκπαίδευσης:

$$P(t) = p(O_1 O_2 O_3 \dots O_t | \lambda^{(t)}) = \prod_{k=1}^t p(O_k | \lambda^{(t)})$$

Αν η σχετική διαφορά των πιθανοτήτων για δύο διαδοχικά βήματα γίνει μικρότερη από ένα κατώφλι τότε ο αλγόριθμος τερματίζει:

$$|P(t) - P(t-1)|/P(t-1) < Thr$$

Αν κατά την διαδικασία εκτίμησης των κρυφών καταστάσεων των παραδειγμάτων παρατηρηθεί ότι η ακολουθίες αυτές είναι ακριβώς ίδιες για όλα τα παραδείγματα με τις εκτιμήσεις που έγιναν στο προηγούμενο βήμα, τότε ο αλγόριθμος τερματίζει διότι η στατιστική εκτίμηση των παραμέτρων του μοντέλου που ακολουθεί θα μας δώσει ακριβώς τις ίδιες παραμέτρους.

Μπορούμε επίσης να τερματίσουμε το επαναληπτικό τμήμα του αλγόριθμου όταν παρατηρήσουμε ότι η σχετική μεταβολή των εκτιμώμενων πιθανοτήτων του μοντέλου είναι αριθμητικά πολύ μικρή.

Παράδειγμα. Μπορείτε να απαντήσετε στα ερωτήματα που σας τέθηκαν στην προηγούμενη άσκηση όταν δεν γνωρίζετε ποιά είναι η πραγματική ακολουθία δυαδικών αριθμών που στέλνει

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

κάθε δορυφόρος;

Να λάβετε υπόψη ότι από τον πρώτο δορυφόρο έχουν ληφθεί τα εξής σήματα αναγνώρισης ταυτότητας:

{ 00111, 1000111, 0011 }

ενώ από τον δεύτερο δορυφόρο παρατηρήθηκαν τα εξής σήματα:

{ 010101, 011101, 0101 }

Παράδειγμα. Έστω ότι θέλουμε να προσομοιώσουμε φυσικό φαινόμενο με ένα κρυφό μοντέλο Markov. Η ακολουθία κρυφών καταστάσεων περιγράφεται από τέσσερα σύμβολα, ενώ η ακολουθία των παρατηρήσεων αποτελείται από δέκα σύμβολα.

1. Πόσες είναι οι ελεύθερες παράμετροι που πρέπει να υπολογίσουμε κατά την διαδικασία της εκπαίδευσης.
2. Αν θέλουμε να υλοποιήσουμε ένα σύστημα ταξινόμησης δέκα κατηγοριών, κάθε μία των οποίων περιγράφεται στοχαστικά από ένα κρυμμένο μοντέλο Markov με τα χαρακτηριστικά που δώσαμε ήδη, υπολογίστε τον χρόνο που απαιτείται για να ταξινομηθεί μία ακολουθία 20 παρατηρήσεων με μικρουπολογιστικό σύστημα το οποίο έχει τα ακόλουθα χαρακτηριστικά:

Πολλαπλασιασμός 500ns, πρόσθεση 180ns. Θεωρείστε αμελητέο τον χρόνο ανάθεσης τιμής και σύγκρισης.

Το συνεχές κρυμμένο μοντέλο Markov

Σε πολλές περιπτώσεις η ακολουθία παρατηρήσεων που λαμβάνουμε σε διακεκριμένα διαστήματα δεν έχει πεπερασμένο πεδίο τιμών, αλλά συνεχές. Σε αυτή την περίπτωση μπορούμε να προσομοιώσουμε πρότυπα που μεταβάλλονται στο χρόνο με την βοήθεια του συνεχούς κρυμμένου μοντέλου Markov.

Η μοναδική διαφοροποίηση που παρατηρείται σε σχέση με το διακριτό μοντέλο εντοπίζεται στις πιθανότητες παρατήρησης δεδομένης της κατάστασης στο κρυφό επίπεδο. Στο διακριτό μοντέλο αυτές οι πιθανότητες είναι διακριτές ενώ στο συνεχές μοντέλο χρησιμοποιούμε μία συνάρτηση πυκνότητας πιθανότητας για κάθε κρυφή κατάσταση.

Οι αλγόριθμοι με τους οποίους υπολογίζουμε την πιθανότητα μίας ακολουθίας παρατηρήσεων (forward procedure) ή την ακολουθία των κρυφών καταστάσεων δεδομένης της ακολουθίας παρατηρήσεων (Viterbi) παραμένουν οι ίδιοι και στην

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

περίπτωση του συνεχούς μοντέλου.

Στην θέση των πιθανοτήτων των καταστάσεων στο επίπεδο των παρατηρήσεων τοποθετούμε την αριθμητική τιμή της συνάρτησης πυκνότητας πιθανότητας για την θέση που προσδιορίζεται από το διάλυμα της παρατήρησης.

Η διαφοροποίηση που υπάρχει κατά την εκπαίδευση του συνεχούς μοντέλου επικεντρώνεται στον τρόπο με τον οποίο υπολογίζονται οι σταθερές παράμετροι της συνάρτησης πυκνότητας πιθανότητας.

Για τις περισσότερες συναρτήσεις πυκνότητας πιθανότητας γνωρίζουμε τις σχέσεις που συνδέουν τις ροπές με τις σταθερές παραμέτρους των. Έτσι λοιπόν από τα παραδείγματα υπολογίζουμε αρχικά τις στατιστικές εκτιμήσεις των ροπών χαμηλής τάξης, και στην συνέχεια υπολογίζουμε τις σταθερές παραμέτρους των συναρτήσεων πυκνότητας πιθανότητας.

Παράδειγμα. Ας δούμε σαν παράδειγμα τον επίγειο δορυφορικό σταθμό που λαμβάνει σήματα από δύο δορυφόρους. Γνωρίζουμε ότι ο πρώτος δορυφόρος στέλνει την ακολουθία 000111 και ο δεύτερος 010101.

Ο δέκτης στον επίγειο σταθμό λαμβάνει αναλογικές τιμές δίνοντας για τον μεν πρώτο δορυφόρο τις ακολουθίες παρατηρήσεων

$$\{(0.2, 0.4, 0.9, 0.8, 0.6, 0.7), (1.1, 0.1, 0, 0.2, 0.5, 0.9), (0.1, 0.9, 0.2, 0.4, 0.7, 1.0)\}$$

για τον δεύτερο δορυφόρο και εγράφησαν τα σήματα

$$\{(0.9, 0.8, 0.3, 1.1, 0, 1), (0.2, 0.9, 1, 0.6, 0.5, 1), (1, 0.8, 0.1, 1, 0.4, 0.7)\}.$$

1. Μπορείτε να προσομοιώσετε την διαδικασία εκπομπής και λήψης σήματος με ένα συνεχές κρυμμένο μοντέλο Markov υποθέτοντας ότι οι συναρτήσεις πυκνότητας πιθανότητας ακολουθούν κανονική κατανομή;
2. Υπολογίστε τις παραμέτρους των κρυφών μοντέλων Markov για το σήμα ταυτότητας που εκπέμπει κάθε δορυφόρος από τις παρατηρήσεις του επίγειου σταθμού.
3. Ποιός δορυφόρος εκπέμπει αν λάβουμε σήμα ταυτότητας το 000000;
4. Ποιός δορυφόρος εκπέμπει αν λάβουμε σήμα ταυτότητας το 000;

Παράδειγμα. Ποιά θα είναι η απάντηση στο τρίτο ερώτημα του προηγούμενου παραδείγματος αν υποθέσετε ότι η συνάρτηση πυκνότητας πιθανότητας ακολουθεί την τριγωνική κατανομή;

Παράδειγμα. Θέλουμε να κατασκευάσουμε ένα αυτόματο ιατρικό διαγνωστικό σύστημα για μία ασθένεια X για την οποία γνωρίζουμε από παρατηρήσεις ότι παρουσιάζει την ακόλουθη συμπεριφορά συμπτωμάτων. Συνήθως στις τρεις πρώτες ημέρες εκδήλωσης της ασθένειας παρουσιάζεται πυρετός με μέση τιμή 38.5 και διασπορά 2. Κατά μέσο όρο τις επόμενες δύο ημέρες παρατηρείται μία ύφεση με μέση τιμή θερμοκρασίας σώματος 37.1 και διασπορά 0.9. Στο τελευταίο στάδιο της ασθένειας η μέση θερμοκρασία αυξάνεται σε 38.1 με διασπορά 0.3.

Γνωρίζουμε ότι η μέση θερμοκρασία ενός υγιούς ανθρώπου είναι 36.5 με διασπορά 0.5.

Στο νοσοκομείο εισήχθη άνθρωπος για τον οποίο γνωρίζουμε ότι τις τελευταίες επτά ημέρες παρουσίασε την ακόλουθη ακολουθία θερμοκρασιών:

37.3, 37.3, 37.0, 36.5, 37.8, 36.9, 37.8, 37.1

Αν υποθέσετε ότι οι ακολουθίες θερμοκρασιών για υγιείς και ασθενείς ανθρώπους μπορούν να προσομοιωθούν με δύο διαφορετικά κρυμμένα μοντέλα Markov, μπορείτε να διαγνώσετε αν ο άνθρωπος είναι υγιής ή ασθενής;

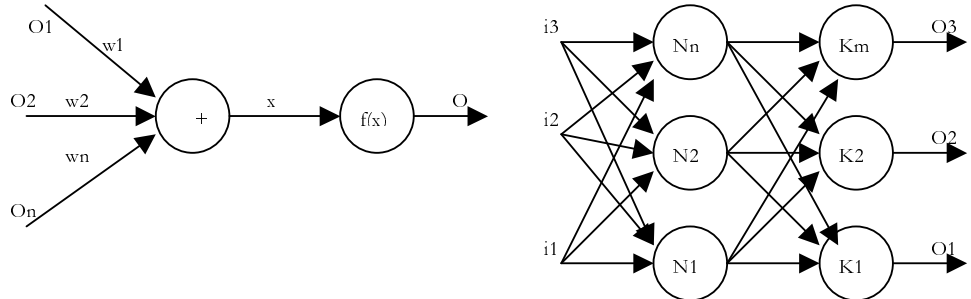
Νευρωνικά δίκτυα και ταξινόμηση χρονοσειρών

Ένας από τους πρώτους ερευνητές που συνέλαβε την ιδέα ότι οι εγκεφαλικές λειτουργίες εκτελούνται από στοιχειώδεις υπολογιστικές μονάδες που ονομάζονται νευρώνες είναι ο Ramon y Cajal (1911). Πειράματα που έγιναν στην φυσιολογία των νευρικών κυττάρων έδειξαν ότι η λειτουργία τους μπορεί, σε απλοποιημένη μορφή, να προσομοιωθεί με μη γραμμικούς τελεστές που δέχονται σήματα από έναν πεπερασμένο αριθμό εισόδων και διαθέτουν μονάχα μία έξοδο. Η εγκεφαλική ουσία των έμβριων όντων αποτελείται από εκατομμύρια διασυνδεδεμένων νευρώνων τα οποία αποτελούν ένα πολύπλοκο και ισχυρά μη γραμμικό νευρωνικό δίκτυο. Έχει αποδειχθεί ότι η γνώση του κόσμου συσσωρεύεται στην μνήμη του δικτύου. Η τοπολογία του διαφέρει σημαντικά στα έμβρια όντα και είναι αυτή η οποία δίνει και τα διαφορετικά χαρακτηριστικά αντίληψης των ερεθισμάτων του περιβάλλοντος. Έχουν παρατηρηθεί ποσοτικές αλλά και ποιοτικές διαφοροποιήσεις στο είδος των συνδέσεων, τον μηχανισμό επεξεργασίας σημάτων και το πλήθος των εισόδων των νευρικών κυττάρων. Γιαυτό τον λόγο και τα νευρικά κύτταρα έχουν ομαδοποιηθεί σε έναν μικρό αριθμό κατηγοριών, ανάλογα με τους μηχανισμούς λειτουργίας των.

Η υπάρχουσα τεχνολογία δεν επιτρέπει την προσομοίωση των πολύπλοκων επεξεργασιών που εκτελούνται στα νευρικά κύτταρα. Έχει αποδειχθεί όμως ότι απλές προσομοιώσεις των νευρικών κυττάρων, που περιγράφουν τα βασικά τους γνωρίσματα, δίνουν εντυπωσιακά αποτελέσματα σε εφαρμογές ταξινόμησης προτύπων, προσεγγίσεων απόκρισης συστημάτων, προβλήματα αυτόματου ελέγχου, κ.α..

Κάθε νευρωνικό δίκτυο αποτελείται από διασυνδεδεμένες υπολογιστικές μονάδες που ονομάζονται νευρώνες. Κάθε νευρώνας μετασχηματίζει το δiάνυσμα εισόδου του δίνοντας μία μονάχα έξοδο η οποία συνδέεται με εισόδους άλλων νευρώνων όπως φαίνεται και στο σχήμα 4.1.

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II



Σχήμα 4.2 Νευρώνας και τυπική μορφή νευρωνικού δικτύου

Μελέτες που αναφέρονται στην φυσιολογία του νευρικού κυττάρου έδειξαν ότι η έξοδος αποτελεί έναν μη γραμμικό μετασχηματισμό των εισόδων του. Οι προσομοιωτές που θα περιγραφούν αποτελούν προσέγγιση των μετασχηματισμών που πραγματοποιούνται στο εσωτερικό των φυσικών νευρικών κυττάρων.

Οι σημαντικότερες προσομοιώσεις είναι υπολογιστικά απλές και αποδίδουν τους βασικούς μηχανισμούς ενεργοποίησης και εκπαίδευσης του φυσικού μοντέλου ικανοποιώντας τα χαρακτηριστικά του γνωρίσματα που περιγράφηκαν στο προηγούμενο εδάφιο.

Οι περισσότερες προσομοιώσεις που έχουν προταθεί βασίστηκαν σε μελέτες του τρόπου λειτουργίας των νευρικών κυττάρων στο φυσικό τους περιβάλλον. Από την απλοποίηση των παρατηρήσεων της φυσιολογίας των νευρικών κυττάρων προέκυψαν τα διάφορα μαθηματικά μοντέλα. Το σημαντικότερο και το πλέον απλό μοντέλο περιγράφει την συμπεριφορά ενός νευρικού κυττάρου με δύο τελεστές, έναν γραμμικό και έναν μη γραμμικό που είναι συνδεδεμένοι σε σειρά, όπως φαίνεται και στο σχήμα.

Ο γραμμικός τελεστής παριστάνεται σαν το εσωτερικό γινόμενο του διανύσματος της εισόδου με το διάνυσμα της μνήμης του νευρώνα:

$$x = ow = \sum_{i=1}^N o_i w_i$$

Ο μη γραμμικός τελεστής είναι μία μη γραμμική συνάρτηση της εξόδου του γραμμικού τελεστή:

$$o = f(\mathbf{x})$$

Η έξοδος του δικτύου συμβολίζεται με το ίδιο γράμμα με το οποίο συμβολίζεται και η είσοδος του νευρώνα διότι η έξοδος του νευρώνα είναι είσοδος σε άλλους νευρώνες. Η συνάρτηση του μη γραμμικού τελεστή ακολουθεί τα χαρακτηριστικά των

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

σιγμοειδών (sigmoid) συναρτήσεων τα οποία είναι τα ακόλουθα:

Είναι αύξουσα συνάρτηση

$$\forall x_1, x_2 \in \mathcal{R}, x_1 > x_2 \rightarrow f(x_1) \geq f(x_2)$$

Έχει πεπερασμένα απειροστικά όρια

$$\lim_{x \rightarrow +\infty} f(x) = a, a \in \mathcal{R} - \{-\infty, +\infty\}$$

$$\lim_{x \rightarrow -\infty} f(x) = b, b \in \mathcal{R} - \{-\infty, +\infty\}$$

Έχει πεδίο ορισμού το σύνολο των πραγματικών αριθμών και έχουν φραγμένο πεδίο τιμών. Η ιδιότητα αυτή προκύπτει σαν αποτέλεσμα των δύο προηγούμενων περιορισμών.

$$f: \mathcal{R} \rightarrow [a, b]$$

Οι πλέον διαδεδομένοι σε πρακτικές εφαρμογές γραμμικοί νευρωνικοί τελεστές είναι οι ακόλουθοι:

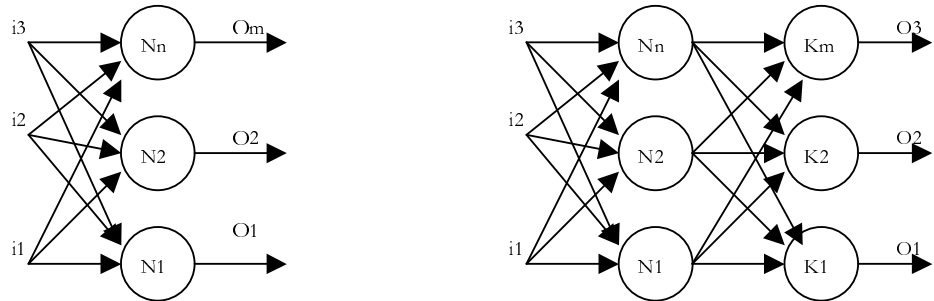
Εκθετική σιγμοειδής. Η συνάρτηση αυτή αποτελεί τον πλέον διαδεδομένο μη γραμμικό νευρωνικό τελεστή:

$$f(x) = \frac{1}{e^{-ax+b} + c}, a, b \in \mathcal{R}^+$$

Υπερβολική εφαπτομένη.

$$f(x) = \frac{e^{ax} - e^{-ax}}{e^{ax} + e^{-ax}}, a \in \mathcal{R}^+$$

Παρόλο που δεν υπάρχει περιορισμός για τον τρόπο με τον οποίο οργανώνονται οι συνάψεις των νευρώνων, υπάρχουν μερικές ομάδες δομών οι οποίες έχουν μελετηθεί εκτενέστερα και είναι αυτές οι οποίες χρησιμοποιούνται συχνότερα σε διάφορες εφαρμογές.



Σχήμα 4.3 Νευρωνικό δίκτυο ενός επιπέδου και Νευρωνικό δίκτυο πολλών επιπέδων

Νευρωνικά δίκτυα ενός επιπέδου. Αποτελεί την πλέον απλή περίπτωση οργάνωσης ενός νευρωνικού δικτύου. Οι εισόδοι κάθε νευρώνα συνδέονται με τις αντίστοιχες εισόδους του δικτύου και η έξοδος κάθε νευρώνα αποτελεί και έξοδο του δικτύου. Σε μερικές περιπτώσεις θεωρούμε ότι ο κάθε νευρώνας έχει και μία επιπλέον είσοδο η οποία συνδέεται με μία είσοδο σταθερής στάθμης.

Τα δίκτυα ενός επιπέδου νευρώνων χρησιμοποιούνται συνήθως σε απλά προβλήματα διότι έχουν δύο σοβαρά μειονεκτήματα.

Αν υποθέσουμε ότι το σύστημα που θέλουμε να προσομοιώσουμε έχει N εισόδους και M εξόδους, τότε στην καλύτερη των περιπτώσεων (όταν κάθε νευρώνας έχει συνάψεις που συνδέουν όλες τις εισόδους) το σύστημα προσομοίωσης έχει $(N+1)M$ βαθμούς ελευθερίας, γεγονός που περιορίζει την ικανότητα του δικτύου να προσομοιώνει πολύπλοκες διανυσματικές συναρτήσεις. Ο περιορισμός αυτός στην πράξη μπορεί να ξεπεραστεί επεκτείνοντας με τεχνητό τρόπο το διάνυσμα της εισόδου οπότε αυξάνουμε ταυτόχρονα και τους βαθμούς ελευθερίας του συστήματος προσομοίωσης. Η επέκταση αυτή μπορεί να επιτευχθεί όταν συνδυάσουμε με γραμμικό ή μη γραμμικό τρόπο τις αρχικές εισόδους του συστήματος.

Ο δεύτερος περιορισμός έχει να κάνει με τις περιορισμένες δυνατότητες του δικτύου στην προσομοίωση των μη γραμμικών χαρακτηριστικών της συνάρτησης που θέλουμε να προσομοιώσουμε. Συγκεκριμένα μπορούμε να δούμε ότι κάθε έξοδος μπορεί να εκφραστεί σαν συνάρτηση των δεδομένων εισόδου στην εξής μορφή:

$$o_m = f\left(\sum_i x_i w_{mi}\right)$$

Η έξοδος o_m μπορεί να προσομοιώσει την πραγματική έξοδο του συστήματος μέσω μιας περιορισμένων δυνατοτήτων συνάρτησης. Κλασσικό πρόβλημα των περιορισμένων δυνατοτήτων των νευρωνικών δικτύων ενός επιπέδου στην ταξινόμηση προτύπων αποτελεί και η αδυναμία σωστής ταξινόμησης μη γραμμικά διαχωρίσιμων

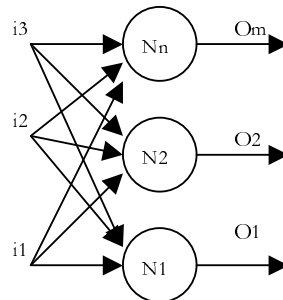
προτύπων.

Πολυεπίπεδα νευρωνικά δίκτυα. Προσθέτοντας ένα ή περισσότερα κρυφά επίπεδα μπορούμε να αυξήσουμε απεριόριστα τους βαθμούς ελευθερίας του νευρωνικού δικτύου. Συνήθης τακτική είναι οι νευρώνες κάθε επιπέδου να συνδέονται με τις εξόδους των νευρώνων που βρίσκονται στο προηγούμενο επίπεδο. Όταν ο νευρώνας συνδέεται με όλους τους νευρώνες του προηγούμενου επιπέδου η σύνδεση ονομάζεται *πλήρης* διαφορετικά ονομάζεται *μερική*.

Κατά την διάρκεια εκπαίδευσης του νευρωνικού δικτύου αν κάποιος συντελεστής βαρύτητας σύναψης πάρει την τιμή μηδέν, τότε το γεγονός αυτό ισοδυναμεί με την διακοπή σύνδεσης των νευρώνων διότι δεν μεταφέρεται πλέον πληροφορία από αυτή την σύναψη. Στην πράξη λοιπόν το δίκτυο μπορεί να εμφανίζει αρχιτεκτονική πλήρους σύνδεσης αλλά κατά την διάρκεια εκπαίδευσης κάποιες συνδέσεις μπορεί να αποκοπούν, γεγονός όχι σπάνιο σε πρακτικές εφαρμογές.

Στην ταξινόμηση προτύπων έχει αποδειχθεί ότι ένα δίκτυο που αποτελείται από νευρώνες perceptron δύο επιπέδων (ενός κρυφού και του επιπέδου εξόδου) μπορεί να ταξινομήσει σωστά γραμμικά και μη γραμμικά διαχωρίσιμες κατηγορίες προτύπων, ενώ ένα δίκτυο τριών επιπέδων νευρώνων μπορεί να ταξινομήσει σωστά κατηγορίες προτύπων μη γραμμικά διαχωρίσιμες που επιπλέον βρίσκονται σε διαφορετικές νησίδες στο χώρο των μετρήσεων.

Πολυεπίπεδα νευρωνικά δίκτυα έχουν χρησιμοποιηθεί με επιτυχία στην σχεδίαση πολύπλοκων συστημάτων ταξινόμησης προτύπων και καλύπτουν την πλειονότητα των εφαρμογών ταξινόμησης προτύπων με νευρωνικά δίκτυα. Αν στην είσοδο του δικτύου το οποίο έχει μόνο μία είσοδο, τοποθετηθούν δειγματοληπτημένα σήματα διακριτού χρόνου κατά αύξουσα χρονική σειρά τότε το δίκτυο λειτουργεί σαν ένα μη γραμμικό φίλτρο FIR (σχήμα 4).

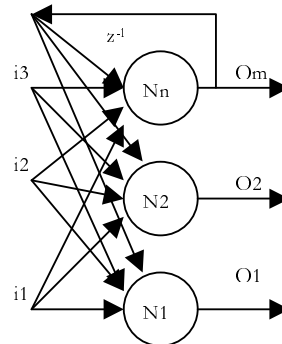


Σχήμα 4.4 Νευρωνικό δίκτυο πολλών επιπέδων σαν μη γραμμικό φίλτρο FIR

Ανατροφοδοτούμενα νευρωνικά δίκτυα. Όταν υπάρχει έστω και μία διαδρομή μέσω της οποίας, ξεκινώντας από έναν νευρώνα και μέσω των συνάψεων και κατά την φορά ενεργοποίησης του δικτύου μπορούμε να επανέλθουμε στον νευρώνα εκκίνησης, τότε το νευρωνικό δίκτυο θα λέγεται ανατροφοδοτούμενο (recurrent).

Τυπικό παράδειγμα ανατροφοδοτούμενου νευρωνικού δικτύου δίνεται στο σχήμα που ακολουθεί.

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ ΙΙ



Σχήμα 4.5 Τυπική μορφή ανατροφοδοτούμενου νευρωνικού δικτύου

Η αρχιτεκτονική αυτή προσδίδει στο νευρωνικό δίκτυο κάποια ιδιαίτερα χαρακτηριστικά. Οι χρονικές καθυστερήσεις δίνουν την δυνατότητα στο δίκτυο να προσομοιώσει χρονικά μεταβαλλόμενα πρότυπα.

Το πλέον διαδεδομένο νευρωνικό δίκτυο συνεχών τιμών εισόδου και εξόδου είναι το πολυεπίπεδο perceptron. Μια τυπική μορφή του δικτύου φαίνεται στο σχήμα. Αποτελείται από νευρώνες του τύπου perceptron οι οποίοι διαθέτουν έναν γραμμικό και έναν μη-γραμμικό τελεστή συνδεδεμένους σε σειρά:

$$y = \sum_{i=1}^N w_i x_i$$

$$o = f(y)$$

Το δίκτυο αποτελείται από περισσότερες των δύο ομάδων νευρώνων perceptron που ονομάζονται και επίπεδα του δικτύου. Ένα ή περισσότερα κρυφά επίπεδα επεξεργάζονται τα δεδομένα εισόδου και το επίπεδο εξόδου περιέχει τους νευρώνες, η έξοδος των οποίων είναι ταυτόχρονα και έξοδος του δικτύου.

Η συνήθης σύνδεση των νευρώνων perceptron ενός επιπέδου περιλαμβάνει συνάψεις που συνδέουν κάθε νευρώνα με όλους τους νευρώνες του προηγούμενου επιπέδου. Η σύνδεση αυτή ονομάζεται πλήρης (full connected) και χρησιμοποιείται σε περιπτώσεις στις οποίες δεν μπορούμε να ορίσουμε εκ των προτέρων την ακριβή τοπολογία του δικτύου. Η γενική αυτή τοπολογία καλύπτει και την πλειονότητα των εφαρμογών.

Η ευρεία διάδοση του πολυεπίπεδου δικτύου perceptron οφείλεται σε δύο κυρίως λόγους:

1. Η ισχυρή πολυπλοκότητα και μη γραμμικότητα του δικτύου μας παρέχει την δυνατότητα να προσομοιώσουμε με ικανοποιητική ακρίβεια συνεχείς μη

γραμμικές διανυσματικές συναρτήσεις.

2. Ο αλγόριθμος εκπαίδευσης LMS δίνει την δυνατότητα υπολογισμού των συντελεστών βαρύτητας των συνάψεων από παραδείγματα. Η δυνατότητα αυτή μας επιτρέπει να χρησιμοποιήσουμε το δίκτυο σε πρακτικές εφαρμογές.

Πράγματι από το 1985, όταν ο Werdos στην διδακτορική του διατριβή παρουσίασε την αρχική μορφή ενός αλγόριθμου εκπαίδευσης που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα της πραγματικής από την επιθυμητή έξοδο του δικτύου και που αργότερα πήρε το όνομα οπισθοδρομική διάδοση του σφάλματος (backpropagation of error), μεγάλος αριθμός εργασιών που αφορούσαν το πολυεπίπεδο δίκτυο perceptron παρουσιάστηκαν σε εφαρμογές ταξινόμησης προτύπων, αυτόματο έλεγχο, προσομοίωση συναρτήσεων, κ.ο.κ.

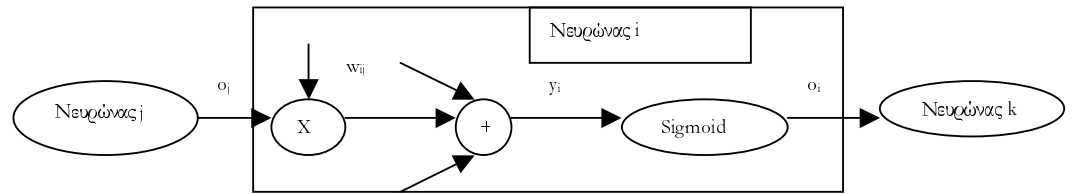
Ο αλγόριθμος οπισθοδρομικής διάδοσης του σφάλματος

Όπως είδαμε και στο προηγούμενο εδάφιο κατά την σχεδίαση ενός γραμμικού φίλτρου Wiener η μέθοδος της κατευθυνόμενης εκπαίδευσης LMS έχει σαν σκοπό την εύρεση των συντελεστών βαρύτητας των συνάψεων που ελαχιστοποιούν το μέσο τετραγωνικό σφάλμα της πραγματικής από την επιθυμητή έξοδο του δικτύου.

Οι ίδιες αρχές ακολουθούνται και στον αλγόριθμο οπισθοδρομικής διάδοσης του σφάλματος με την διαφορά ότι στο πολυεπίπεδο δίκτυο perceptron δεν μπορούμε να πετύχουμε αναλυτική λύση. Γιαντό τον λόγο καταφεύγουμε στην μέθοδο της διαδοχικής προσέγγισης των συντελεστών βαρύτητας των συνάψεων κατά την διεύθυνση της αρνητικής τιμής της πρώτης παραγώγου του σφάλματος (steepest descent - βαθύτατη κάθοδος). Η μέθοδος αυτή που είναι η πλέον διαδεδομένη τεχνική προσδιορισμού της μνήμης πολύπλοκων νευρωνικών ονομάζεται αλγόριθμος οπισθοδρομικής διάδοσης του σφάλματος.

Ο αλγόριθμος οφείλει την ονομασία του στο γεγονός ότι ενώ κατά την ενεργοποίηση του νευρωνικού δικτύου οι υπολογισμοί ξεκινούν από τους νευρώνες εισόδου και κατευθύνονται διαδοχικά προς τους νευρώνες του εξόδου, υπολογίζοντας την τιμή της εξόδου των νευρώνων στα κρυφά επίπεδα, η εκπαίδευση εκτελεί την αντίστροφη διαδικασία. Μετά την ενεργοποίηση του δικτύου αρχίζει η διαδικασία επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων αρχικά στις συνάψεις που βρίσκονται στους νευρώνες εξόδου και στην συνέχεια ακολουθεί ο επαναπροσδιορισμός των συντελεστών βαρύτητας προς τους νευρώνες εισόδου.

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II



Σχήμα 4.6. Διάταξη νευρώνων στον αλγόριθμο οπισθοδρομικής διάδοσης του σφάλματος

Στην γενική περίπτωση ένα πολυεπίπεδο δίκτυο perceptron αποτελείται από έναν αριθμό κρυφών επιπέδων και το επίπεδο εξόδου που περιέχει M νευρώνες. Στην κατευθυνόμενη εκπαίδευση τα παραδείγματα αποτελούνται από ζεύγη διανυσμάτων εισόδου-εξόδου. Αν η είσοδος του δικτύου αποτελείται από N εισόδους τότε τα παραδείγματα είναι διατεταγμένες δυάδες διανυσμάτων (\mathbf{a}, \mathbf{b}) , $\mathbf{a} \in \mathcal{R}^N$, $\mathbf{b} \in \mathcal{R}^M$.

Το σφάλμα εκτίμησης για τυχαία είσοδο \mathbf{a} είναι το τετράγωνο των διαφορών της πραγματικής εξόδου του δικτύου \mathbf{o} με το αντίστοιχο διάνυσμα των αναμενόμενων τιμών \mathbf{b} :

$$\Sigma\phi = 0.5(\mathbf{b} - \mathbf{o})^T (\mathbf{b} - \mathbf{o}) = 0.5 \sum_{i=1}^M (b_i - o_i)^2$$

Δεδομένου ότι συνήθως έχουμε στην διάθεσή μας Q παραδείγματα, το αναμενόμενο σφάλμα υπολογίζεται από την σχέση:

$$E[\Sigma\phi] = 0.5 E\left[\sum_{i=1}^M (b_i - o_i)^2 \right] = 0.5 \sum_{i=1}^M E[(b_i - o_i)^2] = \frac{1}{2Q} \sum_{i=1}^M \sum_{j=1}^Q (b_{ij} - o_{ij})^2$$

Είναι γνωστό ότι οι προσπάθειες των ερευνητών για την αναζήτηση μεθόδου αναλυτικού υπολογισμού των συντελεστών βαρύτητας των συνάψεων, μίας λύσης παρόμοιας με αυτή των φίλτρων Wiener, έχουν καταλήξει σε αποτυχία. Γιαυτό τον λόγο καταφεύγουμε στην μέθοδο της διαδοχικής προσέγγισης των συντελεστών βαρύτητας των συνάψεων κατά την διεύθυνση της αρνητικής τιμής της πρώτης παραγώγου του σφάλματος. Η μέθοδος αυτή επαναπροσδιορίζει την αριθμητική τιμή του συντελεστή βαρύτητας των συνάψεων σύμφωνα με την ακόλουθη σχέση:

$$\delta i = -n \frac{\partial \Sigma\phi}{\partial w_{ij}}$$

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

όπου n είναι μικρός θετικός πραγματικός αριθμός που ονομάζεται ρυθμός εκπαίδευσης του αλγόριθμου. Η εξίσωση μπορεί να εκφραστεί σαν συνάρτηση της εξόδου και της εσωτερικής κατάστασης του νευρώνα:

$$\Delta w_{ij} = -n \frac{\partial \Sigma \phi}{\partial y_i} \frac{\partial y_i}{\partial w_{ij}}$$

Αν ορίσουμε σαν δ_i την αρνητική τιμή της παραγώγου του σφάλματος ως προς την εσωτερική κατάσταση του νευρώνα y_i (εσωτερική κατάσταση του νευρώνα είναι η έξοδος του γραμμικού του τελεστή), τότε:

$$\delta_i = -\frac{\partial \Sigma \phi}{\partial y_i} = -\frac{\partial \Sigma \phi}{\partial o_i} \frac{\partial o_i}{\partial y_i}$$

και

$$\Delta w_{ij} = n \delta_i \frac{\partial y_i}{\partial w_{ij}} = n \delta_i \frac{\partial}{\partial w_{ij}} \sum_k w_{ik} o_k = n \delta_i o_j$$

Για να πετύχουμε μία αναλυτική έκφραση της συνάρτησης επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων χωρίζουμε τους νευρώνες σε δύο κατηγορίες.

Νευρώνες εξόδου

Η αρνητική τιμή της παραγώγου του σφάλματος ως προς την εσωτερική κατάσταση του νευρώνα δ_i γίνεται:

$$\delta_i = -\frac{\partial \Sigma \phi}{\partial o_i} \frac{\partial o_i}{\partial y_i} = -(-(b_i - o_i)) \frac{\partial f(y_i)}{\partial y_i}$$

Ο τελευταίος πολλαπλασιαστικός όρος δεν είναι τίποτα άλλο από την πρώτη παράγωγο του μη γραμμικού τελεστή του νευρώνα. Η συνάρτηση επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων για τους νευρώνες εξόδου γίνεται:

$$\Delta w_{ij} = n \delta_i o_j = n(b_i - o_i) o_j \frac{\partial f(y_i)}{\partial y_i}$$

Κρυφοί νευρώνες

Η περίπτωση των κρυφών νευρώνων είναι ελαφρώς πολυπλοκότερη. Αν υποθέσουμε ότι στο αμέσως υψηλότερο επίπεδο βρίσκονται K νευρώνες, τότε η αρνητική τιμή της παραγώγου του σφάλματος ως προς την εσωτερική κατάσταση του νευρώνα δ_i γίνεται:

$$\delta_i = -\frac{\partial \Sigma \phi}{\partial o_i} \frac{\partial o_i}{\partial y_i} = -\frac{\partial o_i}{\partial y_i} \sum_{k=1}^K \frac{\partial \Sigma \phi}{\partial y_k} \frac{\partial y_k}{\partial o_i} = \frac{\partial o_i}{\partial y_i} \sum_{k=1}^K \left(-\frac{\partial \Sigma \phi}{\partial y_k}\right) \frac{\partial}{\partial o_i} \sum_{m=1}^I w_{km} o_m$$

όπου I είναι ο αριθμός των νευρώνων του επιπέδου που μελετάμε. Μετά την απλοποίηση των παραπάνω εκφράσεων έχουμε:

$$\delta_i = \frac{\partial o_i}{\partial y_i} \sum_{k=1}^K \delta_k w_{ki}$$

Βλέπουμε λοιπόν ότι ο όρος δ του νευρώνα i εξαρτάται από τους αντίστοιχους όρους δ των νευρώνων που βρίσκονται στο αμέσως υψηλότερο επίπεδο και τους συντελεστές βαρύτητας των συνάψεων που συνδέουν την έξοδο του νευρώνα. Προκύπτει μάλιστα σαν το εσωτερικό γινόμενο των αντίστοιχων διανυσμάτων τους. Η συνάρτηση επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων για τους κρυφούς νευρώνες γίνεται:

$$\Delta w_{ij} = n \delta_i o_j = n o_j \frac{\partial o_i}{\partial y_i} \sum_{k=1}^K \delta_k w_{ki}$$

Επειδή ο υπολογισμός των συντελεστών δ είναι αναδρομικός, από τους νευρώνες υψηλότερων επιπέδων προς χαμηλότερα επίπεδα γιατί και η διαδικασία επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων στον πολυεπίπεδο δίκτυο perceptron ακολουθεί τα εξής βήματα:

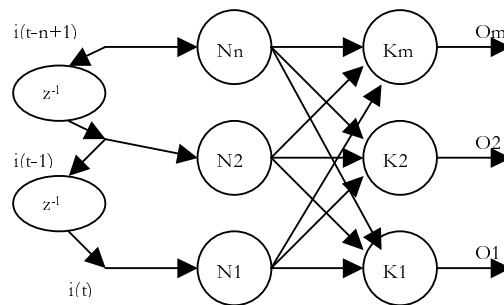
- **Τοποθέτηση αρχικών τιμών.** Ορίζουμε τον ρυθμό εκπαίδευσης n και τοποθετούμε τυχαίες αρχικές τιμές στους συντελεστές βαρύτητας των συνάψεων.
- **Ενεργοποίηση του δικτύου.** Τοποθετούμε στην είσοδο του δικτύου τυχαίο πρότυπο από τα παραδείγματα και στην συνέχεια υπολογίζουμε από χαμηλότερο προς υψηλότερο επίπεδο την έξοδο των νευρώνων κάθε επιπέδου.
- **Υπολογισμός των συντελεστών δ .** Αντιστρέφοντας την φορά των υπολογισμών υπολογίζουμε τους συντελεστές δ από υψηλότερα σε χαμηλότερα επίπεδα.

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

- **Επαναπροσδιορισμός των συντελεστών βαρύτητας των συνάψεων.** Έχοντας υπολογίσει τους συντελεστές δ και την αριθμητική τιμή των εξόδων των κόμβων μπορούμε να επαναπροσδιορίσουμε τους συντελεστές βαρύτητας των συνάψεων όλων των νευρώνων του δικτύου.
- **Έλεγχος σύγκλισης.** Υπάρχουν διάφορα κριτήρια ελέγχου της σύγκλισης του αλγόριθμου. Οι συνήθεις έλεγχοι που γίνονται είναι οι εξής: Υπολογίζεται η συνολική απόλυτη μεταβολή των συντελεστών βαρύτητας σε σχέση με έναν πολύ μικρό θετικό αριθμό, το κριτήριο σύγκλισης. Υπολογίζεται ο μέσος ρυθμός μεταβολής του σφάλματος στα παραδείγματα. Αν η μεταβολή αυτή είναι πολύ μικρή τότε ο αλγόριθμος τερματίζει. Αν ο μέσος ρυθμός μεταβολής των συντελεστών δ για όλα τα παραδείγματα είναι πολύ μικρός, αποδεικνύεται ότι η αντίστοιχη μεταβολή των συντελεστών βαρύτητας των συνάψεων είναι επίσης πολύ μικρή.
- **Επανάληψη του αλγόριθμου.** Τοποθετούμε νέο παράδειγμα στην είσοδο του δικτύου και επαναλαμβάνουμε την διαδικασία από το δεύτερο βήμα.

Το δίκτυο χρονικών καθυστερήσεων

Η αναγνώριση χρονικά μεταβαλλόμενων προτύπων με πεδίο τιμών στο N-διάστατο χώρο των πραγματικών αριθμών μπορεί να πραγματοποιηθεί με τροποποιημένες μορφές του βασικού πολυεπίπεδου δικτύου νευρώνων του τύπου perceptron. Θα περιγράψουμε τα πλέον απλά και διαδεδομένα δίκτυα με πρώτο το νευρωνικό δίκτυο χρονικών καθυστερήσεων (Time-delay neural network).



Σχήμα 4.7 Το νευρωνικό δίκτυο χρονικών καθυστερήσεων

Η δομή του δικτύου χρονικών καθυστερήσεων δίνεται στο σχήμα 4.7. Η βασική του διαφοροποίηση από το πολυεπίπεδο δίκτυο perceptron εντοπίζεται στον τρόπο τοποθέτησης των διανυσμάτων εισόδου. Όπως φαίνεται και από το σχήμα, τα διανύσματα που προσδιορίζονται από την βαθμίδα εξαγωγής παραμέτρων κατά την χρονική στιγμή t τοποθετούνται στην είσοδο του νευρωνικού δικτύου, ενώ ταυτόχρονα οι προηγούμενες τιμές του διανύσματος της εισόδου μετατοπίζονται κατά

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

μία θέση κατά την φορά που δείχνουν οι χρονικές καθυστερήσεις. Με την τεχνική αυτή υπάρχει πάντα μία χρονική ακολουθία K διανυσμάτων στην είσοδο του νευρωνικού δικτύου. Το δίκτυο ενεργοποιείται κάθε χρονική στιγμή που εισέρχεται στο δίκτυο ένα νέο διάνυσμα. Το γεγονός αυτό έχει σαν άμεση συνέπεια την παραγωγή μίας αντίστοιχης ακολουθίας διανυσμάτων εξόδου του δικτύου.

Η ίδια η τοπολογία του δικτύου περιορίζει σημαντικά τις δυνατότητες του δικτύου στην αναγνώριση χρονοσειρών. Η πληροφορία που επεξεργάζεται το δίκτυο αναφέρεται πάντα σε K διαδοχικά χρονικά διανύσματα. Συνεπώς το δίκτυο χρονικών καθυστερήσεων μπορεί να ταξινομεί πρότυπα τα οποία έχουν πολύ μικρή διασπορά παραμέτρων στο πεδίο του χρόνου, ή με άλλα λόγια το φαινόμενο που προσπαθούμε να αναγνωρίσουμε πρέπει να έχει πάντα την ίδια χρονική εξέλιξη. Το δίκτυο μπορεί να καλύψει με επιτυχία την διασπορά των προτύπων στο διανυσματικό χώρο των παραμέτρων του.

Ιδιαίτερα επιτυχημένη είναι η εφαρμογή του δικτύου χρονικών καθυστερήσεων σε συστήματα αναγνώρισης φωνημάτων, διότι οι κατηγορίες ταξινόμησης (τα φωνήματα) ακολουθούν γενικά δύο χαρακτηριστικά μεταβάσεων στο χρόνο τα οποία ικανοποιούν γενικά τους περιορισμούς που τέθηκαν από την μορφή της τοπολογίας του τα φωνήεντα και πολλά σύμφωνα χαρακτηρίζονται από συγκεκριμένες τιμές που λαμβάνει το διάνυσμα των παραμέτρων κατά την βραχύχρονη ανάλυση του σήματος ομιλίας και Τα φωνήματα που χαρακτηρίζονται Από μία βραχύχρονης διάρκειας εκρηκτική συμπεριφορά όπως τα 'κ', 'π' 'τ'. Σε αυτή την κατηγορία φωνημάτων έχει μετρηθεί ότι η διασπορά χρονικής εξέλιξης του φωνήματος είναι σχεδόν μηδενική.

Η εκπαίδευση του δικτύου πραγματοποιείται με τον αλγόριθμο οπισθοδρομικής διάδοσης του σφάλματος. Συγκεκριμένα για κάθε ακολουθία διανυσμάτων τίθενται οι επιθυμητές τιμές εξόδου και στην συνέχεια εφαρμόζεται ο αλγόριθμος οπισθοδρομικής διάδοσης ο οποίος επαναπροσδιορίζει τις τιμές των συντελεστών βαρύτητας των νευρώνων. Το πρόβλημα που πρέπει να λυθεί είναι πως ορίζουμε τις ιδεατές τιμές για να έχουμε μία επιτυχημένη εκπαίδευση. Η αντιμετώπιση δεν είναι μοναδική και διαφέρει στις εφαρμογές.

Στα πλαίσια του μαθήματος θα εξετάσουμε δύο γενικές περιπτώσεις εκπαίδευσης νευρωνικών δικτύων χρονικής καθυστέρησης σε προβλήματα ταξινόμησης προτύπων.

Έστω ότι για την κατηγορία ω_i διαθέτουμε M_i χρονοσειρές παραμετρικών διανυσμάτων ίσου μήκους K διανυσμάτων:

$$V_i = \{ \mathbf{x}_{i,1}[t], \mathbf{x}_{i,2}[t], \dots, \mathbf{x}_{i,M_i}[t] \}$$

Αν γνωρίζουμε ότι η κατηγορία αντικειμένων περιγράφεται πλήρως από K χρονικά ακολουθιακά διανύσματα τότε, μπορούμε να κατασκευάσουμε και να εκπαιδεύσουμε δύο είδη νευρωνικών δικτύων.

Το πρώτο είδος δικτύου περιλαμβάνει ένα ανεξάρτητο δίκτυο χρονικής καθυστέρησης K διανυσμάτων και αριθμό εξόδων ίσο με τον αριθμό των κατηγοριών που το σύστημα αναγνωρίζει. Το δίκτυο εκπαιδεύεται τοποθετώντας στην είσοδο του δικτύου ένα διάνυσμα εκπαίδευσης, έστω το $x_{ij}[t]$, και θεωρώντας σαν ιδεατή έξοδο

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

του δικτύου διάνυσμα που έχει την μεγαλύτερη αριθμητική τιμή στον νευρώνα εξόδου που χρησιμοποιείται για να ταξινομήσει τα διανύσματα της ω_i κατηγορίας και μικρή αριθμητική τιμή σε όλες τις άλλες εξόδους του δικτύου. Η ταξινόμηση πραγματοποιείται με την τοποθέτηση της άγνωστης χρονικής ακολουθίας K διανυσμάτων στην είσοδο του δικτύου, την ενεργοποίησή του, η οποία θα αποδώσει τιμές στους νευρώνες της εξόδου και την ταξινόμηση στην έξοδο που παρουσιάζει την μεγαλύτερη αριθμητική τιμή.

Το δεύτερο είδος δικτύου περιλαμβάνει αριθμό δικτύων ίδιο με τον αριθμό των κατηγοριών. Κάθε δίκτυο έχει μόνο μία έξοδο. Τα δίκτυα εκπαιδεύονται τοποθετώντας στην είσοδο ένα διάνυσμα εκπαίδευσης, έστω το $x_{i,j}[t]$. Στο δίκτυο, το οποίο θα χρησιμοποιηθεί για να αναγνωρίζει τα πρότυπα της ω_i κατηγορίας τοποθετούμε σαν ιδεατή τιμή μία μεγάλη αριθμητική τιμή, ενώ σε όλα τα άλλα δίκτυα τοποθετούμε σαν ιδεατή έξοδο μία πολύ μικρή αριθμητική τιμή.

Η ταξινόμηση πραγματοποιείται με την τοποθέτηση της άγνωστης χρονικής ακολουθίας K διανυσμάτων στην είσοδο όλων των δικτύων, την ενεργοποίησή τους η οποία θα αποδώσει τιμές στον νευρώνα της εξόδου και την ταξινόμηση στο δίκτυο εκείνο που παρουσιάζει την μεγαλύτερη αριθμητική τιμή στον νευρώνα εξόδου.

Η απλή περίπτωση εφαρμογής του νευρωνικού δικτύου χρονικών καθυστερήσεων για πρότυπα στα οποία γνωρίζουμε την χρονική στιγμή εμφάνισης και την διάρκεια τους εμφανίζεται σπάνια σε πρακτικές εφαρμογές.

Συχνότερα εμφανίζεται το εξής διπλό πρόβλημα

1. εύρεσης του χρονικού διαστήματος στο οποίο εκδηλώνεται η εμφάνιση ενός γεγονότος, και
2. την ταξινόμηση του πρότυπου σε κατηγορία γεγονότων.

Η λύση αυτού του προβλήματος επιβάλλει και την διερεύνηση της δυνατότητας εμφάνισης ταυτόχρονων ή αμοιβαία αποκλειόμενων γεγονότων. Με άλλα λόγια πρέπει να μελετήσουμε την συμπεριφορά των προτύπων που αναγνωρίζουμε. Υπάρχουν προβλήματα κατά τα οποία παρακολουθούμε την χρονική εξέλιξη ενός μόνο γεγονότος (όπως το παράδειγμα που δώσαμε για το οποίο υποθέσαμε σιωπηλά ότι ο δέκτης λαμβάνει το σήμα δήλωσης ταυτότητας από ένα μόνο δορυφόρο, ταυτόχρονη εκπομπή δεν είναι αποδεκτό γεγονός). Αντιθέτως σε ένα σύστημα αυτόματης ιατρικής διάγνωσης δεν μπορούμε να αποκλείσουμε την ταυτόχρονη εξέλιξη πολλαπλών ασθενειών. Συνεπώς σε αυτή την κατηγορία προβλημάτων πρέπει να τροποποιήσουμε το είδος της ταξινόμησης που πραγματοποιεί το νευρωνικό δίκτυο που θα κατασκευάσουμε.

Ιδιαίτερα επιτυχημένες εφαρμογές του νευρωνικού δικτύου χρονικών καθυστερήσεων έχουν υλοποιηθεί στην ψηφιακή επεξεργασία σημάτων και σε προβλήματα πρόβλεψης χρονοσειρών.

Παράδειγμα. Περιγράψτε την τοπολογία του νευρωνικού(ών) δικτύων χρονικής καθυστέρησης, τον τρόπο εκπαίδευσης τους και δικαιολογείστε τις απαντήσεις στα ακόλουθα προβλήματα ταξινόμησης προτύπων.

- a. Το πρόβλημα αναγνώρισης της ταυτότητας του δορυφόρου όπως αυτό περιγράφηκε σε προηγούμενο παράδειγμα.
- b. Σύστημα ιατρικής διάγνωσης που επεξεργάζεται τα δεδομένα δέκα φυσικών και πέντε εργαστηριακών εξετάσεων σε καθημερινή βάση και για χρονικό διάστημα δέκα ημερών. Το σύστημα θα αναγνωρίζει τα ακόλουθα νοσήματα: (α) Ηπατίτιδα Β, (β) Ηπατίτιδα C, (γ) Λοιμώδη μονοπυρήνωση, και (δ) ο εξεταζόμενος είναι υγιής.

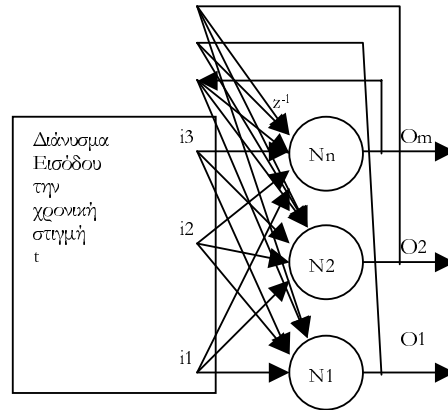
Το δίκτυο Elman

Το μεγάλο μειονέκτημα που παρουσιάζει το νευρωνικό δίκτυο χρονικής καθυστέρησης εντοπίζεται στο γεγονός ότι η εκτίμηση της εξόδου του δικτύου πραγματοποιείται με την επεξεργασία ενός σταθερού χρονικού μήκους "παράθυρο". Γεγονότα τα οποία έχουν διαφορετική χρονική εξέλιξη δεν μπορούν να ταξινομηθούν με ακρίβεια με το νευρωνικό δίκτυο χρονικών καθυστερήσεων.

Λύση σε αυτό το πρόβλημα μπορεί να προκύψει αν στο νευρωνικό δίκτυο τοποθετήσουμε συνάψεις οι οποίες εκτός από τους συντελεστές βαρύτητας περιέχουν και χρονικές καθυστερήσεις. Αν οι συνάψεις αυτές λαμβάνουν πληροφορία από εξόδους νευρώνων του ίδιου ή και ανώτερου επιπέδου, τότε τα νευρωνικά δίκτυα ονομάζονται ανατροφοδοτούμενα (recurrent neural networks).

Ο ορισμός του τρόπου λειτουργίας ανατροφοδοτούμενου δικτύου και οι πιθανές τοπολογίες παρουσιάζουν μεγάλη ποικιλία γιατί και στις επόμενες δύο παραγράφους δίνεται μία αναλυτική περιγραφή των σπουδαιότερων τοπολογιών που απαρτίζονται από νευρώνες τύπου perceptron με τους αντίστοιχους αλγόριθμους εκπαίδευσης.

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ ΙΙ



Σχήμα 4.8 Το νευρωνικό δίκτυο Elman

Το πιο απλό ανατροφοδοτούμενο νευρωνικό δίκτυο προτάθηκε από τον Elman. Η τοπολογία του παρουσιάζεται στο σχήμα 4.8. Στην είσοδο τοποθετείται το διάνυσμα των παραμέτρων που προσδιορίστηκε την χρονική στιγμή t . Το διάνυσμα αυτό μαζί με την αριθμητική τιμή των εξόδων των νευρώνων του πρώτου επιπέδου, (υπολογίστηκαν για το διάνυσμα εισόδου της προηγούμενης χρονικής στιγμής), αποτελούν την πληροφορία που επεξεργάζεται το πολυεπίπεδο (πλήρως διασυνδεδεμένο) δίκτυο perceptron.

Ισχυριζόμαστε ότι το δίκτυο Elman έχει την δυνατότητα να επεξεργαστεί μεταβλητού μήκους γεγονότα. Ας αποδείξουμε τον ισχυρισμό μας.

Αν ορίσουμε σαν $o_i[t]$, $i=1, M$, τις M εξόδους των νευρώνων του πρώτου επιπέδου, και $x_i[t]$, $i=1, N$, το N -διαστάσεων πρότυπο για την χρονική στιγμή t , τότε από την τοπολογία του δικτύου βλέπουμε ότι οι εξοδοί των νευρώνων του πρώτου επιπέδου υπολογίζονται από την ακόλουθη σχέση:

$$o_i[t] = f\left(\sum_{k=1}^N w_{ik} x_k[t] + \sum_{k=1}^M W_{ik} o_k[t-1]\right)$$

Η εξίσωση που προκύπτει είναι αναδρομική. Αν αναλύσουμε τον όρο $o_k[t-1]$ θα πάρουμε την ακόλουθη σχέση

$$o_i[t] = f\left(\sum_{k=1}^N w_{ik} x_k[t] + \sum_{k=1}^M W_{ik} f\left(\sum_{m=1}^N w_{km} x_m[t-1] + \sum_{m=1}^M W_{km} o_m[t-2]\right)\right)$$

Από την εξίσωση βλέπουμε ότι η αριθμητική τιμή των εξόδων των νευρώνων στο

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

πρώτο επίπεδο εξαρτάται από όλες τις τιμές των διανυσμάτων που τέθηκαν στην είσοδο του δικτύου. Η επίδραση αυτή μειώνεται κατά συγκεκριμένο τρόπο με την πάροδο του χρόνου. Ο βαθμός επίδρασης καθορίζεται αποκλειστικά από τους συντελεστές βαρύτητας των συνάψεων που ανατροφοδοτούν το δίκτυο, τους W_{ik} .

Το δίκτυο τύπου Elman βλέπουμε ότι λύνει με απλό τρόπο το πρόβλημα επεξεργασίας μεγάλου μήκους χρονοσειρών. Βέβαια δεν το λύνει με τον καλύτερο δυνατό τρόπο, διότι η θεωρητική εξάρτηση της τιμής της εξόδου επεκτείνεται σε όλα τα διανύσματα που τέθηκαν στην είσοδο, δυσκολεύοντας την προσαρμογή του δικτύου σε προβλήματα στα οποία ζητάμε από το δίκτυο να αναγνωρίσει κατηγορίες γεγονότων που λαμβάνουν χώρα κατά την διάρκεια δειγματοληψίας του χώρου των συμβάντων και έχουν πεπερασμένο μήκος. Επίσης το δίκτυο παρουσιάζει αρκετά μεγάλη ευαισθησία σε γεγονότα που παρουσιάζουν μεγάλη διασπορά στην εξέλιξη. Η ευαισθησία αυτή είναι βέβαια πολύ μικρότερη από την ευαισθησία που παρουσιάζει το δίκτυο χρονικών καθυστερήσεων.

Επίσης καλό είναι να χρησιμοποιούμε την τεχνική εκπαίδευσης πολλαπλών δικτύων, ένα για κάθε κατηγορία γεγονότων, όταν θέλουμε να ταξινομήσουμε κατηγορίες γεγονότων που παρουσιάζουν πολύ μεγάλη διαφορά εξέλιξης στο χρόνο.

Παράδειγμα. Δικαιολογήστε τον ισχυρισμό: όταν θέλουμε να κατασκευάσουμε σύστημα αναγνώρισης προτύπων που περιέχει κατηγορίες γεγονότων που παρουσιάζουν μεγάλη διαφορά στην μέση τους διάρκεια είναι καλύτερο να εκπαιδεύσουμε ανεξάρτητα δίκτυα Elman τα οποία θα έχουν μόνο ένα νευρώνα εξόδου.

Η εκπαίδευση του δικτύου Elman αποτελεί παραλλαγή του αλγόριθμου οπισθοδρομικής διάδοσης του σφάλματος, πιο συγκεκριμένα κατασκευάζουμε τα ιδεατά διανύσματα της εξόδου του δικτύου, επιθυμώντας υψηλή στάθμη εξόδου από τον νευρώνα εξόδου που έχει δεσμευτεί να ταξινομεί τα πρότυπα του γεγονότος που τίθενται στην είσοδο. Η επιθυμητή έξοδος όλων των άλλων νευρώνων τίθενται στην χαμηλότερη τιμή. Ο επαναπροσδιορισμός των συντελεστών βαρύτητας των νευρώνων του δικτύου εκτελείται ως εξής:

Αλγόριθμος

Αρχικοποίηση. Εάν υπάρχουν $M_i, i=1,N$ παραδείγματα χρονικών ακολουθιών προτύπων μπορούμε να κατασκευάσουμε σύστημα ταξινόμησης γεγονότων N κατηγοριών.

Αρχικά επιλέγουμε την κατασκευή ενός δικτύου με πολλές εξόδους ή την κατασκευή πολλών δικτύων, ένα για κάθε κατηγορία γεγονότων. Η επιλογή γίνεται με την παρατήρηση των διαφορών στο χρονικό μήκος που έχουν τα παραδείγματα. Αν παρουσιάζεται σημαντική διαφορά στη μέση χρονική διάρκεια εμφάνισης των γεγονότων στις κατηγορίες, τότε είναι προτιμότερο να κατασκευάσουμε πολλαπλά δίκτυα Elman τα οποία θα αποτελούνται από ένα μόνο νευρώνα εξόδου.

Ορίζουμε την τοπολογία του δικτύου(ων). Οι συντελεστές βαρύτητας των νευρώνων

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

τοποθετούνται σε τυχαίες τιμές κοντά στο μηδέν αλλά όχι σε αριθμητικές τιμές που είναι κατά απόλυτη τιμή μικρότερες ενός κατωφλίου.

Ορίζουμε τις επιθυμητές τιμές εξόδου του δικτύου(ων) σε κάθε ένα από τα παραδείγματα.

Επαναληπτικό τμήμα του αλγόριθμου. Τα επόμενα τρία βήματα του αλγόριθμου επαναλαμβάνονται για έναν αριθμό επαναλήψεων.

- Αρχικοποίηση της εξόδου των νευρώνων του πρώτου επιπέδου. Η έξοδος των νευρώνων του πρώτου επιπέδου τίθεται στο μέσο όρο της ελάχιστης και της μέγιστης τιμής που μπορεί να πάρει ο νευρώνας εξόδου.
- Ενεργοποίηση του δικτύου. Επιλέγεται ένα τυχαίο παράδειγμα και η χρονοσειρά των προτύπων τοποθετείται διαδοχικά στην είσοδο του δικτύου. Για κάθε νέο διάνυσμα που τίθεται στην είσοδο του δικτύου υπολογίζουμε την αριθμητική τιμή όλων των εξόδων των νευρώνων.
- Επαναπροσδιορισμός των συντελεστών βαρύτητας. Επαναπροσδιορίζουμε τους συντελεστές βαρύτητας του δικτύου με την βοήθεια του αλγόριθμου οπισθοδρομικής διάδοσης του σφάλματος, χωρίς να λάβουμε υπόψη την ύπαρξη των συνάψεων που ανατροφοδοτούν το δίκτυο. Οι συντελεστές βαρύτητας αυτών των συνάψεων επαναπροσδιορίζονται θεωρώντας ότι η πληροφορία που λαμβάνουν στην είσοδό του προέρχεται από εξωτερική πηγή.

Παράδειγμα. Περιγράψτε την ιεραρχία του νευρωνικού(ών) δικτύου τύπου Elman, τον τρόπο εκπαίδευσής τους και δικαιολογήστε τις απαντήσεις στα ακόλουθα προβλήματα ταξινόμησης προτύπων

1. *Το πρόβλημα αναγνώρισης της ταυτότητας του δορυφόρου όπως αυτό περιγράφηκε σε προηγούμενο παράδειγμα.*
2. *Σύστημα ιατρικής διάγνωσης που επεξεργάζεται τα δεδομένα δέκα φυσικών και πέντε εργαστηριακών εξετάσεων σε καθημερινή βάση και για χρονικό διάστημα δέκα ημερών. Το σύστημα θα αναγνωρίζει τα ακόλουθα νοσήματα: (α) Ηπατίτιδα Β, (β) Ηπατίτιδα C, (γ) Λοιμώδη μονοπυρήνωση, και (δ) ο εξεταζόμενος είναι υγιής.*

Το δίκτυο Jordan

Το δίκτυο Jordan έχει δομή που μοιάζει με αυτή του δικτύου Elman. Οι βασικές του διαφοροποιήσεις εντοπίζονται σε δύο σημεία.

- Οι συνάψεις που περιέχουν τις χρονικές καθυστερήσεις και ανατροφοδοτούν το δίκτυο παίρνουν πληροφορία από την έξοδο του νευρωνικού δικτύου (από τους νευρώνες που βρίσκονται στο επίπεδο εξόδου). Οι συνάψεις εκτελούν απλά μεταφορά της πληροφορίας από το επίπεδο των νευρώνων εξόδου στο επίπεδο της εισόδου του δικτύου.
- Οι εξοδοί των νευρώνων ανατροφοδότησης όπως και το διάνυσμα του πρότυπου αποτελούν την είσοδο ενός πολυεπίπεδου δικτύου perceptron.

Οι νευρώνες ανατροφοδότησης διαθέτουν επίσης και συνάψεις που περιέχουν καθυστερήσεις και τροφοδοτούν το δίκτυο με πληροφορία που έχει να κάνει με την αριθμητική τιμή της εξόδου του νευρώνα την προηγούμενη χρονική στιγμή:

$$o_i[t] = n o_i[t-1] + O_i[t-1]$$

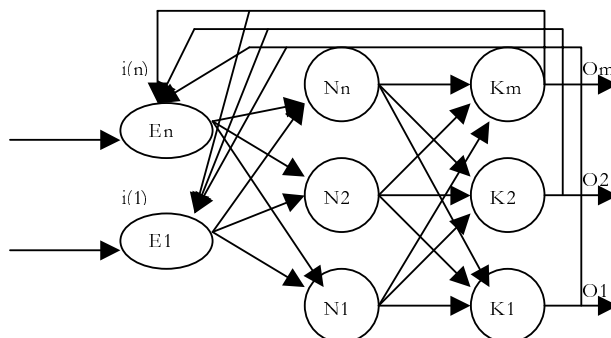
Ο συντελεστής n είναι θετικός αριθμός μικρότερος της μονάδας.

Αν αναλύσουμε την αναδρομική σχέση σε όρους παρελθόντων τιμών θα παρατηρήσουμε ότι η τρέχουσα τιμή της εξόδου των νευρώνων ανατροφοδότησης εξαρτάται από τις τιμές των νευρώνων εξόδου σε παρελθόντα χρόνο και μάλιστα η επίδραση αυτή μειώνεται εκθετικά με τον χρόνο:

$$o_i[t] = o_i[t-1] + n o_i[t-2] + n^2 o_i[t-3] + n^3 o_i[t-4] + \dots$$

Είναι προφανές ότι όσο μεγαλύτερη αριθμητική τιμή δίνουμε στην σταθερά n (πάντα πρέπει να είναι μικρότερη του 1) τόσο μεγαλύτερη θα είναι η επίδραση των παρελθόντος χρόνου τιμών της εξόδου του δικτύου στην αριθμητική τιμή των νευρώνων ανατροφοδότησης.

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II



Σχήμα 4.8. Το νευρωνικό δίκτυο Jordan

Ο υπολογισμός της εξόδου του δικτύου Jordan εκτελείται ως εξής:

Αλγόριθμος

Αρχικοποίηση. Οι έξοδοι ανατροφοδότησης τίθενται σε αριθμητική τιμή που είναι κοντά στο μηδέν π.χ. 0.1.

Επαναληπτικό τμήμα του αλγόριθμου. Τα επόμενα δύο βήματα του αλγόριθμου επαναλαμβάνονται διαδοχικά T φορές. Με T συμβολίζουμε το χρονικό μήκος του πρότυπου. $t \rightarrow 0$.

Ενεργοποίηση του δικτύου. $t \rightarrow t+1$. Το πρότυπο της χρονικής στιγμής t τοποθετείται στην είσοδο του δικτύου. Κατά την διεύθυνση ενεργοποίησης του δικτύου υπολογίζουμε την στάθμη των νευρώνων εξόδου.

Επαναπροσδιορισμός της στάθμης εξόδου των νευρώνων ανατροφοδότησης. Με την βοήθεια της εξίσωσης επαναπροσδιορίζουμε την στάθμη εξόδου των νευρώνων ανατροφοδότησης του δικτύου.

Η εκπαίδευση του δικτύου Jordan είναι ακριβώς ίδια με την εκπαίδευση του δικτύου Elman, γιατί τον λόγο και δεν δίνεται αναλυτική περιγραφή του. Ιδιαίτερη προσοχή πρέπει να δοθεί στην επιλογή της παραμέτρου n, διότι η παράμετρος αυτή επηρεάζει σημαντικά την ταχύτητα σύγκλισης του αλγόριθμου και την ακρίβεια αναγνώρισης των προτύπων. Αν τα πρότυπα παρουσιάζουν μεγάλη χρονική διάρκεια, τότε η παράμετρος αυτή πρέπει να έχει σχετικά μεγάλη αριθμητική τιμή. Οι τυπικές τιμές τις

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

παραμέτρου βρίσκονται στο διάστημα $[0,0.3]$.

Επίσης πολλοί ερευνητές έχουν διαπιστώσει ότι η ταχύτητα σύγκλισης του αλγόριθμου εκπαίδευσης επιταχύνεται όταν, αντί να μεταφέρουμε τις πραγματικές τιμές της εξόδου του δικτύου στους νευρώνες ανατροφοδότησης μεταφέρουμε τις αντίστοιχες ιδεατές τους τιμές. Η βελτίωση που προκύπτει εξαρτάται από την επιλογή και των υπόλοιπων παραμέτρων του αλγόριθμου εκπαίδευσης όπως επίσης εξαρτάται και από το είδος των παραδειγμάτων.

Παράδειγμα. Στα δεδομένα του προηγούμενου παραδείγματος υπολογίστε τον αριθμό των μεταβλητών που πρέπει να εκτιμηθούν από τον αλγόριθμο εκπαίδευσης για τα δίκτυα τύπου Elman και Jordan.

Τι παρατηρείτε;

Ο αλγόριθμος οπισθοδρομικής διάδοσης στο χρόνο

Μία γενική μέθοδος εκπαίδευσης ανατροφοδοτούμενων νευρωνικών δικτύων η οποία δεν περιορίζεται σε μία συγκεκριμένη τοπολογία δικτύου, όπως τα δίκτυα Elman και Jordan, υλοποιείται από τον αλγόριθμο της οπισθοδρομικής διάδοσης στο χρόνο (Back-propagation through Time).

Η εξίσωση που μας δίνει την τιμή εξόδου για κάθε νευρώνα δικτύου που περιέχει ανατροφοδοτήσεις ορίζουμε να είναι η ακόλουθη:

$$o_j[t] = f\left(\sum_i w_{ji} o_i[t] + \sum_k W_{jk} x_k[t]\right)$$

όπου $x_k[t]$, είναι η k εξωτερική είσοδος στο νευρωνικό δίκτυο, και $o_j[t]$ είναι η έξοδος του j νευρώνα κατά την χρονική στιγμή t .

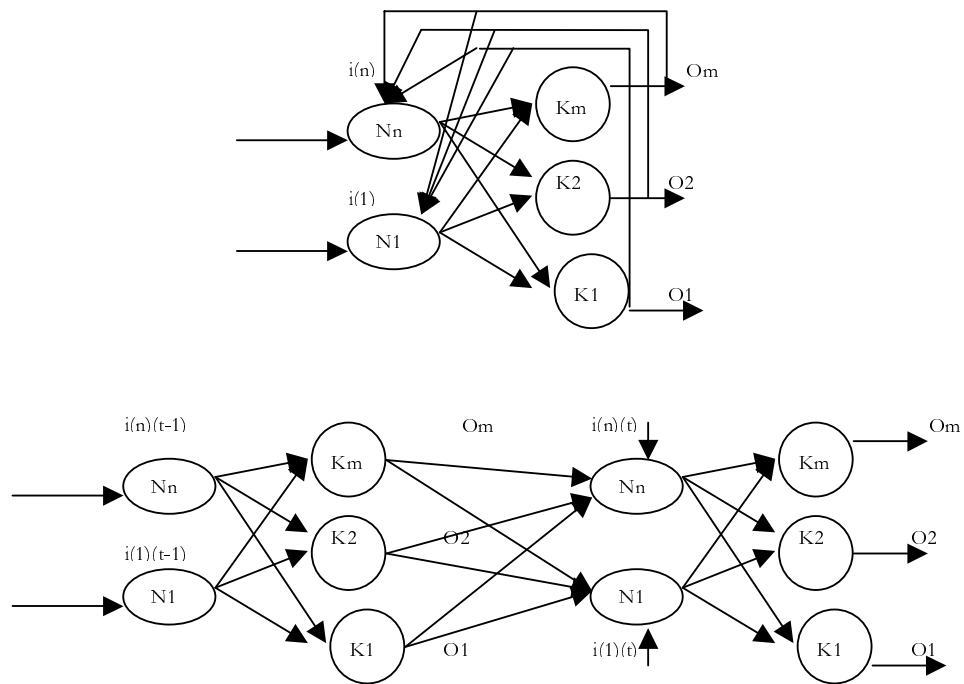
Ας δούμε πως μπορούμε να εκπαιδύσουμε ένα δίκτυο που παρουσιάζει τα χαρακτηριστικά που δώσαμε ή ισοδύναμα με ποιόν αλγόριθμο μπορούμε να υπολογίσουμε τους συντελεστές βαρύτητας του δικτύου έτσι ώστε να ελαχιστοποιήσουμε το σφάλμα της πραγματικής από την ιδεατή έξοδο για ένα πεπερασμένο αριθμό παραδειγμάτων.

Η τεχνική που θα χρησιμοποιήσουμε αποσκοπεί στο να αφαιρέσει την επίδραση της παραμέτρου του χρόνου στο δίκτυο, κατασκευάζοντας ένα ισοδύναμο νευρωνικό δίκτυο το οποίο δεν περιέχει ανατροφοδοτήσεις και χρονικές καθυστερήσεις. Στο δίκτυο που δημιουργείται μπορούμε απλά να εφαρμόσουμε τον αλγόριθμο οπισθοδρομικής διάδοσης του σφάλματος. Η αποσυνέλιξη του δικτύου στο χρόνο μπορεί να εφαρμοστεί από την έξοδο προς την είσοδο του δικτύου είτε κατά την αντίστροφη φορά.

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

Κάθε νευρώνας του δικτύου που αναπτύσσεται στον χρόνο από την ταυτότητα του νευρώνα του ανατροφοδοτούμενου δικτύου και την χρονική στιγμή ενεργοποίησής του.

Πιο συγκεκριμένα ας παρακολουθήσουμε το απλό δίκτυο του σχήματος. Αποτελείται από δύο νευρώνες οι οποίοι δέχονται τα δεδομένα εισόδου και είναι ταυτόχρονα οι εξόδοι του δικτύου. Συνάψεις ανατροφοδότησης προς τον ίδιο και τον άλλον νευρώνα μεταφέρουν με χρονική καθυστέρηση την στάθμη εξόδου στην αντίστοιχη είσοδο του δικτύου.



Σχήμα 4.9 Η αποσυνέλιξη ανατροφοδοτούμενου νευρωνικού δικτύου στο χρόνο

Ας μελετήσουμε την συμπεριφορά του δικτύου για δεδομένα εισόδου μήκους χρόνου ίσο με 3. Το δίκτυο του σχήματος είναι το ισοδύναμο του ανατροφοδοτούμενου δικτύου.

Παρατηρούμε ότι το δίκτυο αυξάνει σε μέγεθος και ότι οι ίδιοι συντελεστές βαρύτητας του δικτύου εμφανίζονται σε όλους τους νευρώνες που προέρχονται από τον ίδιο νευρώνα του ανατροφοδοτούμενου δικτύου, αλλά ενεργοποιούνται σε διαφορετικές χρονικές στιγμές.

Το πρόβλημα που δημιουργείται με τον αλγόριθμο οπισθοδρομικής διάδοσης του σφάλματος εντοπίζεται στο γεγονός ότι οι συντελεστές βαρύτητας του δικτύου μετά τον επαναπροσδιορισμό τους δεν έχουν την ίδια αριθμητική τιμή. Το πρόβλημα αυτό ξεπερνιέται όταν σαν πραγματική τιμή του κάθε συντελεστή βαρύτητας των συνάψεων του δικτύου ανατροφοδότησης οριστεί να είναι η μέση τιμή των

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ II

αντίστοιχων συντελεστών βαρύτητας του δικτύου μετά την εφαρμογή του αλγόριθμου οπισθοδρομικής διάδοσης του σφάλματος.

1. Όσα δεν εμπίπτουν στις παραπάνω κατηγορίες προορίζονται για την εγχώρια αγορά.

Οι αισθητήρες που είναι διαθέσιμοι μπορούν να μετρήσουν σε κάθε πορτοκάλι όγκο, χρώμα, βάρος, πυκνότητα, υγρασία και σκληρότητα της επιφάνειας.

Από το είδος του διαχωρισμού που πρέπει να κάνουμε γίνεται φανερό ότι θα πρέπει να αναζητήσουμε τρόπους έτσι ώστε να διακρίνουμε τα αντικείμενα από τον όγκο, το χρώμα την υγρασία και την σκληρότητα της επιφάνειας.