

Τεχνητά Νευρωνικά Δίκτυα και εφαρμογές στη σύνθεση μουσικής και την αναγνώριση μουσικού συνθέτη

Καλιακάτσος-Παπακώστας Α. Μάξιμος

Διπλωματική Εργασία

Πανεπιστήμιο Πατρών
Σχολή Θετικών Επιστημών
Τμήμα Μαθηματικών
Πάτρα

Επιβλέπων: Καθηγητής Μιχαήλ Ν. Βραχάτης

(Απρίλιος 2009)

Στην Αγγελική, το Θανάση και την Ιωάννα για τη στήριξή τους, τη Λαμπρινή για την υπομονή της και το δάσκαλό μου για την επιμονή του στις κατευθύνσεις που μου έδωσε.

Περιεχόμενα

Ευχαριστίες	1
Περίληψη	3
1 Τεχνητά Νευρωνικά Δίκτυα	5
1.1 Εισαγωγή	5
1.2 Μοντέλα Νευρώνων	6
1.2.1 Τύποι Συναρτήσεων Ενεργοποίησης	8
1.3 Αρχιτεκτονικές Τεχνητών Νευρωνικών Δικτύων	11
1.3.1 Εμπρόσθιας Τροφοδότησης Δίκτυα Ενός Επιπέδου (Single-Layer Feedforward networks)	11
1.3.2 Εμπρόσθιας Τροφοδότησης Δίκτυα Πολλαπλών Επιπέδων (Multilayer Feedforward Networks)	11
1.3.3 Αναδραστικά Τεχνητά Νευρωνικά Δίκτυα (Recurrent Neural Networks (RNN))	12
1.4 Διαδικασίες Μάθησης-Εκπαίδευσης	13
1.4.1 Μάθηση με διόρθωση σφάλματος (Error-Correction Learning)	15
1.4.2 Μάθηση με επίβλεψη (Supervised Learning)	16
1.4.3 Μάθηση χωρίς επίβλεψη (Unsupervised Learning)	17
1.5 Πολυεπίπεδα Εμπρόσθιας Τροφοδότησης ΤΝΔ ή Multilayer Perceptrons (MLP)	17
1.5.1 Ο αλγόριθμος Back-Propagation	19
1.6 Αναδραστικά Νευρωνικά δίκτυα	22
1.6.1 Ο αλγόριθμος Back-Propagation Through Time για τα δίκτυα Elman	23
2 Σύνθεση μουσικής με πρόβλεψη με τη χρήση ΤΝΔ	27
2.1 Εισαγωγή	27
2.2 Προσέγγιση του ζητήματος της αλγοριθμικής μουσικής με πινάκα με- τάβασης	28
2.3 Σύνθεση νότας προς νότα	31
2.4 Αρχιτεκτονική του CONCERT	31
2.4.1 Κανόνες ενεργοποίησης μονάδας	33
2.4.2 Διαδικασία εκπαίδευσης	34
2.5 Αναπαράσταση μουσικών αντικειμένων	36
2.5.1 Αναπαράσταση τονικότητας	37
2.5.2 Αναπαράσταση διάρκειας	41
2.5.3 Αναπαράσταση συγχορδιών	43
2.6 Βασικά αποτελέσματα εξομοιώσεων	44
2.6.1 Επέκταση της κλίμακας C μείζονα	44

2.6.2	Εκπαίδευση στη δομή των διατονικών κλιμάκων	44
2.6.3	Εκπαίδευση σε τυχαίο περίπατο	45
2.6.4	Εκπαίδευση σε διεσπαρμένο τυχαίο περίπατο	45
2.6.5	Εκπαίδευση στις φράσεις με δομή ΑΑΒΑ	46
2.6.6	Συζήτηση των μέχρι τώρα αποτελεσμάτων	46
2.7	Εντοπισμός μουσικής δομής μεγαλύτερης τάξης	47
2.8	Επαναπροσέγγιση των φράσεων ΑΑΒΑ	49
2.9	Πειράματα μεγαλύτερης κλίμακας	49
2.9.1	Σύνθεση μελωδίας στη μορφή συνθέσεων του Bach	50
2.9.2	Σύνθεση μουσικής στη μορφή συνθέσεων ευρωπαϊκών λαϊκών κομματιών	51
2.9.3	Αποκλειστική εκπαίδευση σε δομές υψηλής τάξης: το πείραμα του βαλς	51
2.10	Συζήτηση για την συνολική απόδοση του CONCERT	51
3	Αναγνώριση Μουσικού Συνθέτη με χρήση Τεχνητών Νευρωνικών Δι- κτύων	55
3.1	Εισαγωγή	55
3.2	Εξόρυξη πληροφοριών κειμένου μέσω παρτιτούρας	56
3.2.1	Τι είναι τα MIDI αρχεία	56
3.2.2	Πως χρησιμοποιούμε τα αρχεία MIDI	56
3.2.3	Ουσιώδη στοιχεία μιας παρτιτούρας	57
3.3	Μια πρώτη προσέγγιση στην αναγνώριση συνθέτη από ΤΝΔ	57
3.3.1	Μια απλή προσέγγιση	57
3.3.2	Τοποθέτηση των κομματιών σε ισοδύναμα τονικά μέτρα	58
3.3.3	Τεχνικά ζητήματα	59
3.3.4	Το πρώτο πείραμα	61
3.3.5	Τεχνικά στοιχεία του πρώτου πειράματος	61
3.3.6	Αποτελέσματα του πρώτου πειράματος	63
3.4	Το δεύτερο πείραμα	63
3.4.1	Αποτελέσματα του δεύτερου πειράματος	64
3.5	Συζήτηση	65
	Βιβλιογραφία	67

Κατάλογος Σχημάτων

1.1	Μη-γραμμικό Μοντέλο νευρώνα	7
1.2	Αποτέλεσμα της εισαγωγής της μεροληψίας στον τεχνητό νευρώνα.	8
1.3	Συνάρτηση ενεργοποίησης με κατώφλι.	10
1.4	Τμηματική γραμμική συνάρτηση.	10
1.5	Σιγμοειδής συνάρτηση ενεργοποίησης.	10
1.6	Εμπρόσθιας τροφοδότησης τεχνητό νευρωνικό δίκτυο πολλαπλών επιπέδων με ένα κρυφό επίπεδο και τοπολογία 3-3-2.	12
1.7	Αναδραστικό ΤΝΔ με ένα μόνο επίπεδο	13
1.8	Αναδραστικό ΤΝΔ με κρυφούς νευρώνες	14
1.9	Εκπαίδευση με Επίβλεψη.	16
1.10	Αρχιτεκτονική MLP δικτύου.	19
1.11	Η διαδικασία Back Propagation.	20
1.12	Αρχιτεκτονική SRNN δικτύου.	23
1.13	Ξεδίπλωμα στο χρόνο του παραπάνω απλού αναδραστικού δικτύου σε ένα ισοδύναμο εμπρόσθιας τροφοδότησης. Προσοχή θέλει το ότι τα βάρη V και U χρησιμοποιούνται μια φορά.	24
2.1	Η αρχιτεκτονική του CONCERT	32
2.2	Η ξεδιπλωμένη στον χρόνο αρχιτεκτονική του CONCERT	36
2.3	Η αναπαράσταση PHCCCF	40
2.4	Κλασματική αναπαράσταση νοτών βάσει της μοντελοποίησής μας.	42
2.5	Μοντελοποίηση για την αναπαράσταση της διάρκειας.	42
2.6	Μοντελοποίηση για την αναπαράσταση της διάρκειας.	48
3.1	Πρόχειρη σχηματική αναπαράσταση του χωρισμού των κομματιών για το πρώτο πείραμα σε σύνολο εκπαίδευσης και δοκιμής.	62
3.2	Το σύνολο (a) έχει ομαδοποιημένους στόχους πράγμα που κάνει την εκπαίδευση του δικτύου πιο αναποτελεσματική, αντίθετα με το σύνολο (b).	62

Κατάλογος Πινάκων

2.1	Πίνακας πιθανότητας μετάβασης από τη προσωρινή νότα, που βρίσκεται στη πρώτη γραμμή, στην επόμενη που βρίσκεται στη πρώτη στήλη.	29
2.2	Αναπαράσταση RHCCCF για επιλεγμένες τονικότητες.	38
2.3	Αναπαράσταση της τονικότητας στον χρωματικό κύκλο CC.	38
2.4	Οι δέκα μικρότερες αποστάσεις της αναπαράστασης <i>A1 D2 E2</i>	39
2.5	Οι δέκα μικρότερες αποστάσεις της αναπαράστασης <i>A1 D2 D#2</i> . .	39
2.6	Παράδειγμα συγχορδιών και οι νότες από τις οποίες αποτελούνται. .	43
2.7	Επιδόσεις στις οκτάβες 3 και 4 της C μείζονος διατονικής κλίμακας.	45
2.8	Επιδόσεις των δικτύων με και χωρίς μειωμένη περιγραφή στις φράσεις AABA.	49
3.1	Παράδειγμα πλήθους βαθμίδων μείζονος κλίμακας σε τυχαίο κομμάτι.	59
3.2	Παράδειγμα πυκνότητας βαθμίδων μείζονος κλίμακας στο παραπάνω κομμάτι.	60
3.3	Αποτελέσματα του πρώτου πειράματος για δίκτυα με ένα κρυφό στρώμα. Σκοπός του δικτύου είναι να αναγνωρίσει αν ένα κομμάτι έχει συντεθεί από τον Chopin ή όχι (δηλαδή από τους Beethoven ή Mozart).	64
3.4	Αποτελέσματα του πρώτου πειράματος για δίκτυα με δύο κρυφά στρώματα. Σκοπός του δικτύου είναι να αναγνωρίσει αν ένα κομμάτι έχει συντεθεί από τον Chopin ή όχι (δηλαδή από τους Beethoven ή Mozart).	65
3.5	Αποτελέσματα του δεύτερου πειράματος για δίκτυα με ένα κρυφό στρώμα. Σκοπός του δικτύου είναι να αναγνωρίσει αν ένα κομμάτι έχει συντεθεί από τον Bach ή όχι, δηλαδή από τον Handel	66
3.6	Αποτελέσματα του δεύτερου πειράματος για δίκτυα με δύο κρυφά στρώματα. Σκοπός του δικτύου είναι να αναγνωρίσει αν ένα κομμάτι έχει συντεθεί από τον Bach ή όχι, δηλαδή από τον Handel.	66

Γλωσσάρι

Συντόμογραφια	Περιγραφή	Σελίδα
H/Ψ	Ηλεκτρονικός Υπολογιστής	5
BPTT	Back Propagation Through Time	23
BP	Back Propagation	17
CC	Chroma Circle	37
CF	Circle of Fifths	37
CONCERT	Conectionist composer erutide tunes	27
MIDI	Musical Instument Digital Interface	56
MLP	Multilayer Perceptrons	11, 17
NND	Next Note Destibuted	31
NNL	Next Note Local	31
PH	Pitch Hight	37
RNN	Recurrent Neural Network	12
SRNN	Simple Recurrent Neural Network	22
ΑΤΝΔ	Αναδραστικό Τεχνητό Νευρωνικό Δίκτυο	12
ΤΝΔ	Τεχνητά Νευρωνικά Δίκτυα	5

Ευχαριστίες

Ένας από τους ισχυρότερους λόγους που επέλεξα και επεδίωξα να γίνω μεταπτυχιακός φοιτητής του τομέα υπολογιστών του Μαθηματικού τμήματος του Πανεπιστημίου Πατρών ήταν ευκαιρία που δινότανε να έχω σαν δάσκαλο τον καθηγητή Μ. Ν. Βραχάτη, έναν καθηγητή που από όταν ήμουν προπτυχιακός φοιτητής θαύμαζα για την ευθύτητα, την ευγένεια, τη ψυχραιμία, την οξυδέρκεια, τις πολλές και σε πολλούς τομείς εμπειρίες του και το βάθος γνώσεων που οι εμπειρίες αυτές συνεπάγονται. Κατά τη διάρκεια των σπουδών μου ως μεταπτυχιακός φοιτητής, ο κύριος Βραχάτης σχεδόν με πίεσε να συνδυάσω το πάθος και τις γνώσεις μου για τη μουσική με το πάθος και τις γνώσεις μου για τα μαθηματικά, κάτι που δε θα είχε συμβεί χωρίς αυτόν και θα έχανα ίσως τη μέχρι τώρα σημαντικότερη και πιο περιπετειώδη αναζήτηση της ζωής μου. Οι ευχαριστίες που του οφείλω μέχρι τώρα και πιθανότατα για την υπόλοιπή μου ζωή είναι αμέτρητες.

Η συνεισφορά του διδακτορικού φοιτητή Μιχάλη Επιτροπάκη ήταν σε πολλά επίπεδα καταλυτική. Με τις κορυφαίες προγραμματιστικές του γνώσεις, το μαθηματικό του υπόβαθρο και την αστείρευτη όρεξη του να βοηθήσει αντιμετωπίζοντας τα προβλήματα που αντιμετώπισα όχι μόνο επιδερμικά, αλλά μαθαίνοντας σε βάθος πράγματα που στην ουσία δεν είχε άμεσο κίνητρο να μάθει, με βοήθησε να ξεπεράσω ποικίλες δυσκολίες και με ενέπνευσε στέλνοντάς μου υλικό που θεωρούσε ορθώς ότι θα με ενδιέφερε.

Περιττές αναζητήσεις και κόπο απέτρεψαν τα σχόλια και οι προτάσεις του διδακτορικού φοιτητή Μάνου Οικονομάκη. Επίσης η όρεξη που επέδειξε στο να με βοηθήσει με τη γλώσσα προγραμματισμού PERL με κρατούν υπόχρεο. Θα ήθελα τέλος από τους μεταπτυχιακούς φοιτητές να ευχαριστήσω για την απολαυστική συνεργασία τους Νίκο Κόλλια, Βασίλη Γαλάνη και Σωτήρη Τασουλή καθώς και όσους άλλους με έκανα να αναπτύξω την ανθρώπινη μεριά μου τόσα χρόνια μέσα στο Πανεπιστήμιο Πατρών. Θα ήθελα επίσης να ευχαριστήσω την κυρία Ειρήνη Λαλιώτου που με τις γνώσεις της για τη μουσική έδειξε τους ορθούς δρόμους αναζήτησης.

Από τα βάθη της καρδιάς μου θέλω να ευχαριστήσω τους γονείς μου, Αγγελική και Θανάση, και την αδερφή μου Ιωάννα για την αμέριστη συμπαράστασή τους, τη Λαμπρινή Βασιλείου για την υπομονή και την στοργή της. Τέλος θα ήθελα να ευχαριστήσω τον συνοδοιπόρο μου στις μουσικές αναζητήσεις Σωτήρη Τσακανίκα και τους φίλους μου στην Αθήνα, την Πάτρα και την Κολωνία.

Μάξιμος Καλιακάτσος-Παπακώστας

Πάτρα, 2009.

Περίληψη

Στην παρούσα διπλωματική εργασία μελετάμε την ικανότητα των τεχνητών νευρωνικών δικτύων στη σύνθεση μουσικής και την αναγνώριση μουσικού συνθέτη.

Συγκεκριμένα, στο πρώτο κεφάλαιο κάνουμε μία εισαγωγή στα τεχνητά νευρωνικά δίκτυα και ειδικά σε αυτά που χρησιμοποιούνται στα επόμενα κεφάλαια. Γίνεται αναφορά στα βασικά είδη των ΤΝΔ που υπάρχουν, εμπρόσθιας τροφοδότησης και αναδραστικά και περιγράφονται οι αλγόριθμοι εκπαίδευσής τους.

Εξηγούμε την ικανότητα των αναδραστικών νευρωνικών δικτύων να έχουν δυναμική μνήμη, σε αντίθεση με αυτά που είναι εμπρόσθιας τροφοδότησης πράγμα που τα καθιστά ικανά στην πρόβλεψη χρονοσειρών. Αυτή η ικανότητα των αναδραστικών δικτύων σε συνδυασμό με το γεγονός ότι ένα μουσικό κομμάτι μπορεί να χαρακτηριστεί σαν μία αλληλουχία γεγονότων χρονικής συνοχής (χρονοσειρά) δημιούργησε ένα ερευνητικό ρεύμα προς την κατεύθυνση της σύνθεσης μουσικής με τη χρήση αναδραστικών τεχνητών νευρωνικών δικτύων.

Στο δεύτερο κεφάλαιο κάνουμε μία αναφορά στην αλγοριθμική σύνθεση μουσικής, ιδιαίτερα με χρήση πιθανοτικών πινάκων μετάβασης. Έπειτα ακολουθεί η περιγραφή του CONCERT, ενός αναδραστικού νευρωνικού δικτύου που κατασκευάστηκε για να συνθέτει μουσική με πρόβλεψη νότας προς νότα. Αναλύουμε επίσης την μοντελοποίηση των μουσικών αντικειμένων για την επεξεργασία και αναπαράστασή τους από το CONCERT η οποία βασίζεται σε ψυχοακουστικούς περιορισμούς αντίληψης των μουσικών αντικειμένων από τους ανθρώπους. Εξηγούμε τον τρόπο που εκπαιδεύεται το CONCERT έτσι ώστε να έχει όσο το δυνατόν μεγαλύτερη μνήμη και περιγράφουμε τις επιδόσεις του σε διάφορες δοκιμές που έγιναν, από την εκμάθηση μίας διατονικής κλίμακας μέχρι ενός κομματιού του J.S.Bach.

Παρατηρώντας την ικανότητα του CONCERT να αντιλαμβάνεται τοπικές δομές (μοτίβα, φράσεις) μα όχι καθολικές (μέρη του μουσικού κομματιού) αναφερόμαστε στην τεχνική της περιορισμένης περιγραφής που αποτελεί μια προσπάθεια για εκπαίδευση του δικτύου έτσι ώστε να αντιλαμβάνεται το μουσικό κομμάτι σε μία μεγαλύτερη κλίμακα.

Στο τέλος του δευτέρου κεφαλαίου εξετάζουμε τη συνολική επίδοση του CONCERT και αναλύουμε τις κατευθύνσεις προς τις οποίες θα μπορούσαμε να κινηθούμε για τη βελτίωση των αποτελεσμάτων.

Στο τρίτο κεφάλαιο αναφερόμαστε στην αναγνώριση του συνθέτη ενός μουσικού κομματιού με τη χρήση τεχνητών νευρωνικών δικτύων πάνω στην παρτιτούρα του κομματιού αυτού. Αρχικά γίνεται μια συζήτηση γύρω από το ποια στοιχεία της παρτιτούρας θεωρούμε σημαντικά, ποια από αυτά αυτά μπορούμε και ποια έχει νόημα να μοντελοποιήσουμε έτσι ώστε ένα νευρωνικό δίκτυο να μπορεί να κάνει πρόβλεψη.

Αναλύονται οι τεχνικές λεπτομέρειες των στοιχείων που χρειαζόμαστε για τη μοντελοποίηση μιας παρτιτούρας στον υπολογιστή και στη συνέχεια αναφερόμαστε στα δύο πειράματα που ελέγχουν την ορθότητα και αποτελεσματικότητα της παραπάνω

προσέγγισής μας στο ζήτημα της αναγνώρισης του μουσικού συνθέτη. Το ποια κομμάτια χρησιμοποιήθηκαν και από ποιους συνθέτες δε θα μπορούσε να είναι τυχαίο καθώς πρέπει να ικανοποιούνται διάφορες συνθήκες στατιστικής ομοιομορφίας έτσι ώστε η απάντηση του νευρωνικού δικτύου να είναι όσο το δυνατόν πιο αμερόληπτη. Αυτές οι συνθήκες, καθώς και οι κίνδυνοι που υπάρχουν σε πιθανή παράληψή τους, εξηγούνται πριν τα πειράματα.

Το πρώτο πείραμα πραγματοποιείται την αναγνώριση συνθέτη ενός κομματιού που συντέθηκε από τον Chopin ή όχι (δηλαδή από τους Beethoven ή Mozart) ενώ στο δεύτερο οι εμπλεκόμενοι συνθέτες είναι οι Bach και Handel. Δοκιμάζονται διάφορες αρχιτεκτονικές τεχνητών νευρωνικών δικτύων και μετρούμε την μέση και τη βέλτιστη επίδοσή τους.

Τέλος συζητάμε τα αποτελέσματα των δύο πειραμάτων καθώς και τροποποιήσεις είτε του νευρωνικού δικτύου είτε της μοντελοποίησης που διαλέξαμε για την αναπαράσταση της παρτιτούρας στον υπολογιστή έτσι ώστε να έχουμε καλύτερα αποτελέσματα.

Τεχνητά Νευρωνικά Δίκτυα

1.1 Εισαγωγή

Η ιδέα των Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ) έχει προέλθει από τον τρόπο λειτουργίας των βιολογικών νευρωνικών δικτύων του ανθρώπινου εγκεφάλου. Από την στιγμή που διαπιστώθηκε ότι ο ανθρώπινος εγκέφαλος λειτουργεί εντελώς διαφορετικά από τον τρόπο που λειτουργούν οι Ηλεκτρονικοί Υπολογιστές (Η/Ψ) για να κάνουν υπολογισμούς, ξεκίνησε η προσπάθεια προσομοίωσης των μεθόδων μάθησης και λειτουργίας του από τους Η/Υ.

Η προσπάθεια για την κατανόηση του εγκεφάλου, αρχικά οφείλεται στην πρωτοποριακή εργασία του Ramon y Cajal το 1911, που εισήγαγε την ιδέα των νευρώνων ως τα δομικά χαρακτηριστικά του ανθρώπινου εγκεφάλου. Συνήθως, οι νευρώνες είναι πέντε με έξι τάξεις πιο αργοί από τις λογικές πύλες που κατασκευάζονται από σιλικόνη. Σε ένα chip σιλικόνης συμβαίνουν γεγονότα της τάξης του nanosecond ($10^{-9}s$), αντιθέτως τα γεγονότα στους βιολογικούς νευρώνες συμβαίνουν στην τάξη των millisecond ($10^{-3}s$). Εντούτοις ο εγκέφαλος, αναπληρώνει την σχετικά αργή ταχύτητα επεξεργασίας που έχουν οι νευρώνες του με την ύπαρξη ενός εξαιρετικά μεγάλου αριθμού νευρώνων (νευρικών κυττάρων) με πάρα πολλές συνδέσεις μεταξύ τους. Υπολογίζεται ότι στον ανθρώπινο εγκεφαλικό φλοιό πρέπει να υπάρχουν περί τα 10 δισεκατομμύρια νευρώνες και 60 τρισεκατομμύρια συνάψεις [36].

Ο ανθρώπινος εγκέφαλος είναι ένας ιδιαίτερα πολύπλοκος, μη γραμμικός και παράλληλος ηλεκτρονικός υπολογιστής (σύστημα επεξεργασίας πληροφοριών). Έχει την ικανότητα να οργανώνει τους νευρώνες με τέτοιο τρόπο ώστε να κάνει συγκεκριμένους υπολογισμούς, όπως για παράδειγμα την αναγνώριση προτύπων (pattern recognition), την αντίληψη (perception) και την κίνηση, πολύ πιο γρήγορα από τον γρηγορότερο ηλεκτρονικό υπολογιστή που υπάρχει. Για παράδειγμα αν θεωρήσουμε την ανθρώπινη όραση, η οποία είναι μια εργασία επεξεργασίας πληροφοριών. Ο λειτουργικός ρόλος του οπτικού συστήματος είναι να παρέχει μια αναπαράσταση του περιβάλλοντος γύρω μας και κυρίως να παρέχει τις απαιτούμενες πληροφορίες έτσι ώστε να μπορούμε να αλληλεπιδρούμε με αυτό. Πιο συγκεκριμένα, ο ανθρώπινος εγκέφαλος έχει την ικανότητα να πραγματοποιεί εργασίες αναγνώρισης, π.χ. να αναγνωρίζει ένα γνώριμο πρόσωπο σε μια άγνωστη σκηνή, σε περίπου 100 με 200 ms, ενώ ένας ηλεκτρονικός υπολογιστής μπορεί να χρειάζεται μέρες για να το πραγματοποιήσει.

Ένα άλλο χαρακτηριστικό του ανθρώπινου εγκεφάλου είναι η *πλαστικότητα*, η ικανότητα δηλαδή του νευρικού συστήματος να προσαρμόζεται στο περιβάλλον. Στον εγκέφαλο ενός ενήλικου ανθρώπου, η πλαστικότητα εκφράζεται με δύο τρόπους:

1. με την δημιουργία νέων συνάψεων μεταξύ των νευρώνων και

2. με την τροποποίηση των ήδη υπάρχοντων.

Όπως στον ανθρώπινο εγκέφαλο, έτσι και στα ΤΝΔ η πλαστικότητα είναι απαραίτητη για την λειτουργία των νευρώνων σαν αυτόνομες μονάδες επεξεργασίας πληροφοριών. Στη γενική του μορφή, ένα *νευρωνικό δίκτυο* είναι μια μηχανή που έχει σχεδιαστεί για να μοντελοποιεί τον τρόπο με τον οποίο ο εγκέφαλος εκτελεί μια συγκεκριμένη εργασία. Το δίκτυο συνήθως υλοποιείται με την χρήση ηλεκτρονικών εξαρτημάτων ή προσομοιώνεται με τη χρήση λογισμικού σε έναν ηλεκτρονικό υπολογιστή. Παρακάτω δίνεται ο ορισμός ενός ΤΝΔ που έχει προταθεί στην εργασία [1] από τους Aleksander και Morton το 1990.

Ένα Τεχνητό Νευρωνικό Δίκτυο είναι ένας παράλληλος και κατανεμημένος επεξεργαστής που έχει κατασκευαστεί από απλές μονάδες επεξεργασίας (νευρώνες), που έχει μια φυσική κλίση στο να αποθηκεύει εμπειρική γνώση και έχει την ικανότητα να την χρησιμοποιήσει. Ομοιάζει τον ανθρώπινο εγκέφαλο με δύο τρόπους:

1. Η γνώση λαμβάνεται στο δίκτυο από το περιβάλλον μέσω μιας διαδικασίας εκπαίδευσης
2. Η αποθήκευση της γνώσης γίνεται μέσω των συνάψεων που υπάρχουν στις διασυνδέσεις μεταξύ των νευρώνων και ονομάζονται *βάρη*.

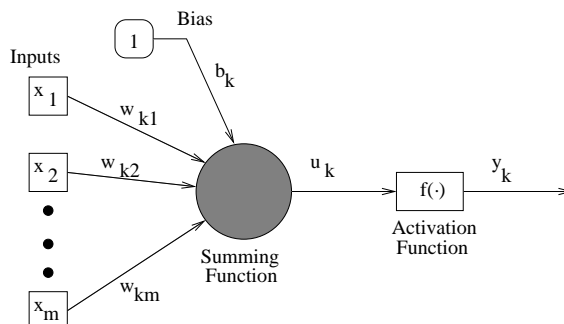
Η διαδικασία που χρησιμοποιείται για την εκπαίδευση ενός ΤΝΔ καλείται *αλγόριθμος εκπαίδευσης*, και ο σκοπός της είναι να μεταβάλει τα βάρη των διασυνδέσεων του δικτύου έτσι ώστε το δίκτυο να παράγει την επιθυμητή έξοδο. Εκτός από την μεταβολή των βαρών ενός ΤΝΔ, το δίκτυο έχει την δυνατότητα να μεταβάλει και την τοπολογία του, όπως συμβαίνει στους νευρώνες του ανθρώπινου εγκεφάλου. δηλαδή, κάποιιοι από τους νευρώνες σταματούν να λειτουργούν και πεθαίνουν ενώ κάποιοι άλλοι δημιουργούν νέες συνδέσεις.

1.2 Μοντέλα Νευρώνων

Ένας νευρώνας είναι μια μονάδα επεξεργασίας πληροφοριών βασική για την λειτουργία του ΤΝΔ. Στο σχήμα 1.1 φαίνεται ένα μοντέλο νευρώνα, το οποίο αποτελεί το βασικό στοιχείο για την σχεδίαση ενός ΤΝΔ. Σε ένα μοντέλο νευρώνα μπορούμε να αναγνωρίσουμε τρία βασικά στοιχεία:

1. Ένα σύνολο από *συνάψεις (διασυνδέσεις)*, η κάθε μια από τις οποίες χαρακτηρίζεται από κάποιο *βάρος*. Συγκεκριμένα, ένα σήμα x_j στην είσοδο μιας σύναψης j που είναι συνδεδεμένη με το νευρώνα k πολλαπλασιάζεται με το βάρος της σύναψης w_{kj} . Οι υποδείκτες του w έχουν την εξής σημασία. Ο πρώτος υποδείκτης αναφέρεται στον εν λόγω νευρώνα και ο δεύτερος στην είσοδο της σύναψης όπου αναφέρεται το βάρος. Το βάρος w_{kj} είναι θετικό όταν η σχετική σύναψη διεγείρει το νευρώνα και αρνητικό όταν η σύναψη είναι ανασταλτική.
2. Έναν *αθροιστή* που αθροίζει τα εισερχόμενα σήματα στον νευρώνα και έχουν πολλαπλασιαστεί με το βάρος της αντίστοιχης σύναψης από την οποία εισήλθαν. Οι διαδικασίες που περιγράφονται εδώ αποτελούν ένα *γραμμικό συνδιαστή*.

3. Μια *συνάρτηση ενεργοποίησης* για τον περιορισμό του μεγέθους της εξόδου ενός νευρώνα. Συνήθως το κανονικοποιημένο εύρος της εξόδου ενός νευρώνα είναι το κλειστό σύνολο $[0, 1]$ ή $[-1, 1]$.



Σχήμα 1.1: Μη-γραμμικό Μοντέλο νευρώνα

Το μοντέλο του νευρώνα που παρουσιάζεται στο σχήμα 1.1 περιλαμβάνει επίσης και ένα εξωτερικό βάρος β_k . Το βάρος β_k έχει ως αποτέλεσμα την αύξηση ή την μείωση της τιμής που δίνει σαν είσοδο το δίκτυο στην συνάρτηση ενεργοποίησης ανάλογα με το αν είναι αρνητικό ή θετικό. Το βάρος αυτό ονομάζεται *μεροληψία* (bias).

Με μαθηματικούς όρους, μπορούμε να περιγράψουμε έναν νευρώνα k από το παρακάτω ζεύγος εξισώσεων:

$$u_k = \sum_{j=1}^m w_{kj} x_j$$

και

$$y_k = \phi(u_k - \beta_k)$$

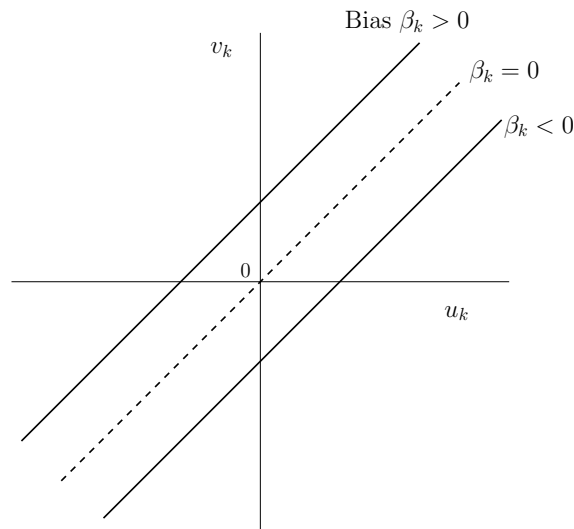
όπου x_1, x_2, \dots, x_m είναι τα εισερχόμενα σήματα, $w_{k1}, w_{k2}, \dots, w_{km}$ είναι τα βάρη των συνάψεων του νευρώνα k , u_k είναι η έξοδος του γραμμικού συνδιαστή, $\phi(\cdot)$ είναι η συνάρτηση ενεργοποίησης και y_k είναι το σήμα που δίνει σαν έξοδο ο νευρώνας k . Η χρήση της μεροληψίας β_k στο μοντέλο του σχήματος 1.1 έχει ως αποτέλεσμα την εφαρμογή ενός μετασχηματισμού της εξόδου u_k με τον γραμμικό συνδιαστή όπως φαίνεται από την παρακάτω εξίσωση:

$$v_k = u_k - \beta_k$$

Συγκεκριμένα, με βάση το αν η μεροληψία β_k είναι θετική ή αρνητική, η σχέση μεταξύ της *δυνατότητας ενεργοποίησης* v_k του νευρώνα k και της εξόδου του γραμμικού συνδιαστή u_k τροποποιείται όπως φαίνεται στο σχήμα 1.2. Σημειώνεται ότι με την εφαρμογή του μετασχηματισμού αυτού η γραφική παράσταση του v_k ως προς το u_k δεν περνά πλέον από την αρχή των αξόνων.

Η μεροληψία β_k είναι μια εξωτερική παράμετρος του τεχνητού νευρώνα k . Μπορούμε να το εντάξουμε στις εξισώσεις που περιγράφουν το νευρώνα όπως πιο πάνω, ή να τις γράψουμε ως εξής:

$$v_k = \sum_{j=0}^m w_{kj} x_j$$



Σχήμα 1.2: Αποτέλεσμα της εισαγωγής της μεροληψίας στον τεχνητό νευρώνα.

και

$$y_k = \phi(v_k)$$

Στην νέα εξίσωση του v_k έχουμε προσθέσει μια νέα σύναψη που έχει σαν είσοδο

$$x_0 = +1$$

και βάρος

$$w_{k0} = \beta_k$$

Με αυτό τον τρόπο δημιουργείται ένα νέο μοντέλο για το νευρώνα. Σε αυτή την περίπτωση ο τρόπος που η μεροληψία επηρεάζει το νευρώνα περιγράφεται από δύο πράγματα: (1) προστίθεται ένα νέο σήμα εισόδου που έχει σταθερά την τιμή +1, (2) προστίθεται ένα νέο βάρος σύναψης που ισούται με το β_k .

1.2.1 Τύποι Συναρτήσεων Ενεργοποίησης

Η συνάρτηση ενεργοποίησης, που συμβολίζεται με $\phi(\cdot)$, ορίζει την έξοδο του νευρώνα σε σχέση με την δυνατότητα ενεργοποίησης στην είσοδό του. Υπάρχουν τρεις βασικοί τύποι συναρτήσεων ενεργοποίησης:

1. Η *συνάρτηση κατώφλι* (Threshold function). Για αυτό το είδος συνάρτησης ενεργοποίησης που περιγράφεται στο σχήμα 1.3 έχουμε

$$\phi(u) = \begin{cases} 1, & \text{αν } u \geq 0 \\ 0, & \text{αν } u \leq 0 \end{cases}$$

αντίστοιχα, η έξοδος του νευρώνα k με την χρήση μιας τέτοιας συνάρτησης εκφράζεται σαν

$$y_k = \begin{cases} 1, & \text{αν } u_k \geq 0 \\ 0, & \text{αν } u_k \leq 0 \end{cases}$$

όπου u_k είναι η δυνατότητα ενεργοποίησης του νευρώνα. Δηλαδή

$$u_k = \sum_{j=1}^m w_{kj} x_j - \beta_k$$

Ένας νευρώνας που χρησιμοποιεί τέτοιου είδους συνάρτηση ενεργοποίησης ακολουθεί το μοντέλο των McCulloch-Pitts. Σε αυτό το μοντέλο η έξοδος του νευρώνα παίρνει την τιμή 1 αν η ολική δυνατότητα ενεργοποίησης του νευρώνα είναι μη αρνητική διαφορετικά παίρνει την τιμή 0.

2. *Τμηματική γραμμική συνάρτηση (Piecewise-Linear Function)*. Για αυτό το είδος συνάρτησης ενεργοποίησης που φαίνεται στο σχήμα 1.4 έχουμε

$$\phi(\cdot) = \begin{cases} 1, & \text{αν } u \geq \frac{1}{2} \\ u, & \text{αν } \frac{1}{2} > u > -\frac{1}{2} \\ 0, & \text{αν } u \leq -\frac{1}{2} \end{cases}$$

όπου ο παράγοντας ενίσχυσης (amplification factor) μέσα στην γραμμική περιοχή της συνάρτησης θεωρείται μονάδα. Το είδος αυτών των συναρτήσεων μπορεί να θεωρηθεί μια προσέγγιση σε μη γραμμικό ενισχυτή. Οι δυο περιπτώσεις παρακάτω είναι ειδικές μορφές της Piecewise-Linear Function:

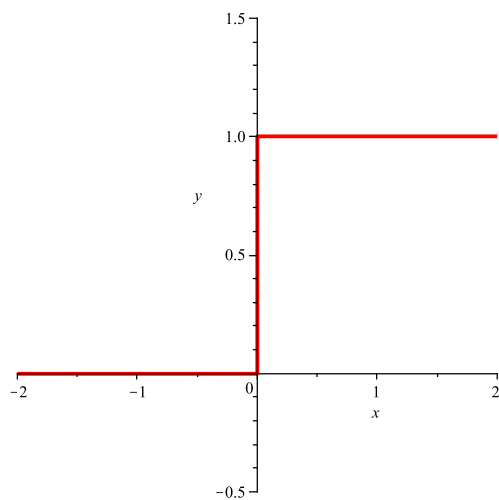
- Ένας γραμμικός συνδιαστής χρησιμοποιείται όταν η γραμμικότητα της περιοχής ενεργοποίησης διατηρείται χωρίς να υπάρχουν κορεσμοί
- Η τμηματική γραμμική μετατρέπεται σε συνάρτηση κατώφλι αν ο παράγοντας ενίσχυσης της γραμμικής περιοχής γίνει απείρως μεγάλος.

3. *Σιγμοειδής συνάρτηση ενεργοποίησης*. Η σιγμοειδής συνάρτηση είναι η πιο ευρέως χρησιμοποιούμενη συνάρτηση ενεργοποίησης για την κατασκευή ΤΝΔ. Ορίζεται ως μια γνησίως αύξουσα συνάρτηση που είναι ομαλή και ασυμπτωτική. Ένα παράδειγμα σιγμοειδούς συνάρτησης είναι η παρακάτω συνάρτηση:

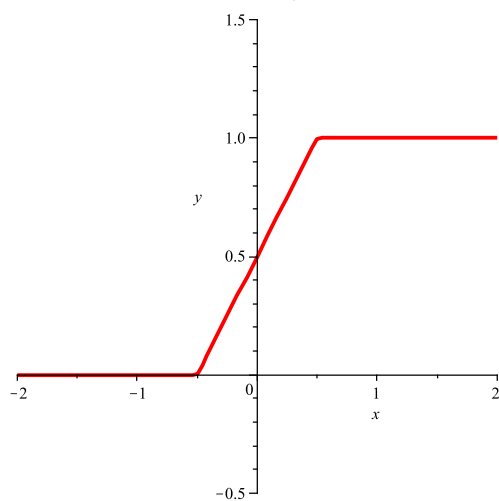
$$\phi(u) = \frac{1}{1 + \exp(-\alpha u)}$$

όπου α είναι η παράμετρος κλίσης της σιγμοειδούς συνάρτησης. Με μεταβολή του α μπορούμε να μεταβάλουμε την κλίση της σιγμοειδούς όπως φαίνεται στο σχήμα 1.5. Συγκεκριμένα η κλίση της σιγμοειδούς στην αρχή των αξόνων ισούται με $\frac{\alpha}{4}$. Στο όριο, καθώς το α τείνει προς το άπειρο, η σιγμοειδής συνάρτηση μετατρέπεται σε συνάρτηση κατώφλι. Αντίθετα με την συνάρτηση κατώφλι που παίρνει τιμές 0 ή 1, η σιγμοειδής συνάρτηση μπορεί να πάρει όλες τις τιμές από το [0,1]. Επίσης η σιγμοειδής συνάρτηση είναι διαφορίσιμη ενώ η συνάρτηση κατώφλι δεν είναι.

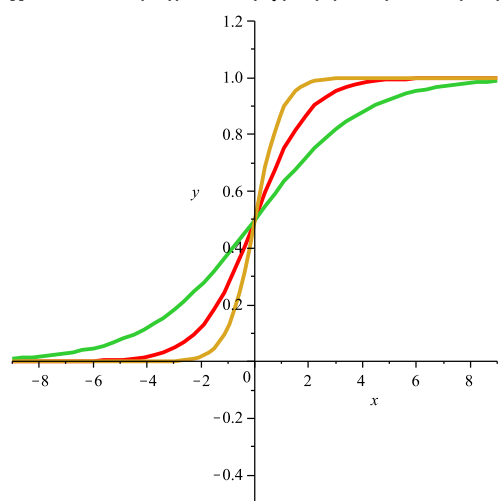
Μέχρι τώρα οι συναρτήσεις που αναφέρθηκαν παίρνουν τιμές από το 0 ως το +1, κάποιες φορές όμως είναι επιθυμητό η συνάρτηση ενεργοποίησης να παίρνει τιμές από το -1 ως το +1. Σε αυτή την περίπτωση η συνάρτηση ενεργοποίησης παίρνει μια αντισυμμετρική μορφή ως προς την αρχή των αξόνων. Συγκεκριμένα η συνάρτηση



Σχήμα 1.3: Συνάρτηση ενεργοποίησης με κατώφλι.



Σχήμα 1.4: Τμηματική γραμμική συνάρτηση.



Σχήμα 1.5: Σιγμοειδής συνάρτηση ενεργοποίησης.

κατώφλι γίνεται

$$\phi(\cdot) = \begin{cases} 1, & \text{αν } u > 0 \\ 0, & \text{αν } u = 0 \\ -1, & \text{αν } u < 0 \end{cases}$$

ενώ για την σιγμοειδή συνάρτηση μπορούμε να πάρουμε την υπερβολική εφαπτομένη που δίνεται από:

$$\phi(u) = \tanh(u)$$

1.3 Αρχιτεκτονικές Τεχνητών Νευρωνικών Δικτύων

Ο τρόπος με τον οποίο οι νευρώνες ενός ΤΝΔ είναι δομημένοι είναι στενά συνδεδεμένος με τον αλγόριθμο εκπαίδευσης που χρησιμοποιείται για το δίκτυο. Γενικά υπάρχουν τρεις βασικές διαφορετικές κλάσεις αρχιτεκτονικών δικτύου.

1.3.1 Εμπρόσθιας Τροφοδότησης Δίκτυα Ενός Επιπέδου (Single-Layer Feedforward networks)

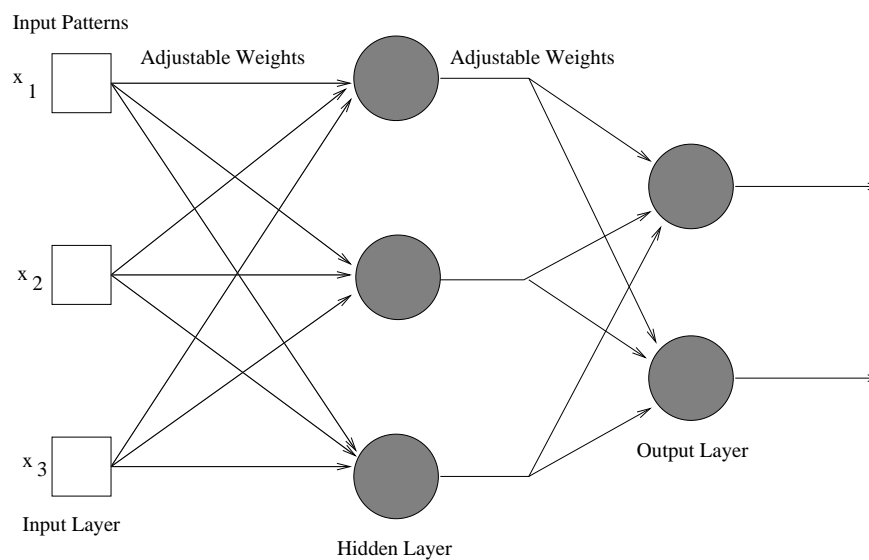
Στα νευρωνικά δίκτυα οι νευρώνες του δικτύου οργανώνονται σε διάφορα επίπεδα. Η πιο απλή μορφή νευρωνικού δικτύου με επίπεδα είναι ένα νευρωνικό δίκτυο με ένα επίπεδο. Στην περίπτωση αυτή έχουμε ένα επίπεδο με κόμβους εισόδου που προβάλλονται σε ένα επίπεδο εξόδου με νευρώνες (όπου γίνονται οι υπολογισμοί), ενώ το αντίθετο δεν ισχύει, δεν μπορούμε δηλαδή να πάμε από τους νευρώνες εξόδου στους κόμβους εισόδου. Σε αυτή την περίπτωση, το δίκτυο είναι ένα αυστηρά εμπρόσθιας τροφοδότησης (feedforward) δίκτυο και καλείται *Εμπρόσθιας Τροφοδότησης Δίκτυο Ενός Επιπέδου*. Με τον όρο 'ένα επίπεδο' εννοούμε το επίπεδο εξόδου που περιέχει και τους νευρώνες όπου γίνονται οι υπολογισμοί. Σημειώνεται ότι δεν υπολογίζεται το επίπεδο εισόδου με τους κόμβους εισόδου γιατί σε αυτό δεν γίνονται καθόλου υπολογισμοί.

1.3.2 Εμπρόσθιας Τροφοδότησης Δίκτυα Πολλαπλών Επιπέδων (Multilayer Feedforward Networks)

Η δεύτερη κλάση εμπρόσθιας τροφοδότησης νευρωνικών δικτύων διαφέρει από την πρώτη στην ύπαρξη ενός ή περισσότερων *κρυφών* επιπέδων, των οποίων οι νευρώνες καλούνται *κρυφοί νευρώνες*. Η λειτουργία των κρυφών νευρώνων είναι να παρεμβάλλονται μεταξύ των κόμβων εισόδου και των νευρώνων εξόδου του δικτύου. Με την προσθήκη ενός ή περισσότερων κρυφών επιπέδων, το δίκτυο έχει την δυνατότητα να προσεγγίζει συναρτήσεις μεγαλύτερης πολυπλοκότητας.

Οι κόμβοι εισόδου στο επίπεδο εισόδου του δικτύου παρέχουν τα στοιχεία των προτύπων εισόδου, με την μορφή διανυσμάτων, που γίνονται εισοδοί στους νευρώνες του δεύτερου επιπέδου, δηλαδή του πρώτου κρυφού επιπέδου. Τα σήματα εξόδου των νευρώνων του δεύτερου κρυφού επιπέδου, γίνονται σήματα εισόδου στο τρίτο επίπεδο και συνεχίζεται κατά αυτό τον τρόπο η ροή των σημάτων μεταξύ των επιπέδων του δικτύου μέχρι να φτάσουν στο επίπεδο εξόδου. Συνήθως, οι νευρώνες σε κάθε επίπεδο του δικτύου, παίρνουν σαν είσοδο, τα σήματα εξόδου από τους νευρώνες του προηγούμενου επιπέδου μόνο. Το σύνολο των σημάτων εξόδου στο επίπεδο εξόδου (τελευταίο επίπεδο του δικτύου), αποτελεί και την απάντηση του δικτύου για

τα δεδομένα που εισάγονται στους κόμβους εισόδου. Στο σχήμα 1.6 φαίνεται ένα εμπρόσθιας τροφοδότησης τεχνητό νευρωνικό δίκτυο πολλαπλών επιπέδων με ένα κρυφό επίπεδο. Για συντομία το δίκτυο του σχήματος αναφέρεται και σαν 3-3-2 δίκτυο αφού έχει 3 κόμβους εισόδου, 3 νευρώνες στο κρυφό επίπεδο και 2 νευρώνες στο επίπεδο εξόδου. Γενικά ένα δίκτυο με m κόμβους εισόδου, h_1 νευρώνες στο πρώτο κρυφό επίπεδο, h_2 νευρώνες στο δεύτερο κρυφό επίπεδο, κ.λπ. h_n νευρώνες στο n -οστό κρυφό επίπεδο και q νευρώνες στο επίπεδο εξόδου, αναφέρεται σαν ένα $m - h_1 - h_2 - \dots - h_n - q$ δίκτυο.



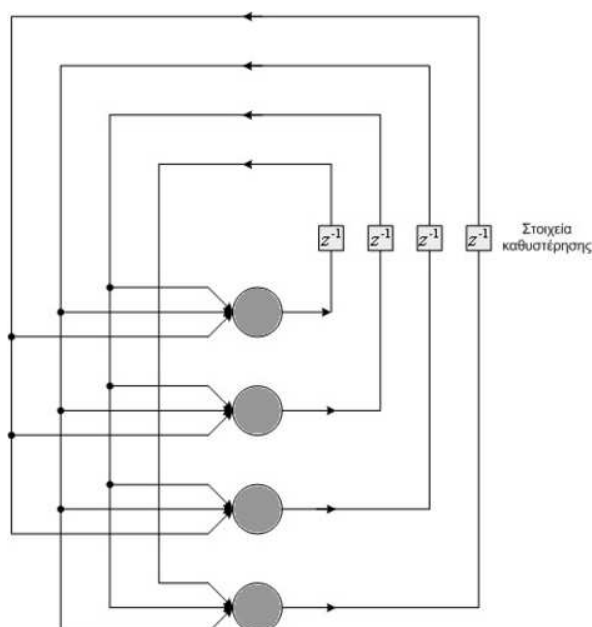
Σχήμα 1.6: Εμπρόσθιας τροφοδότησης τεχνητό νευρωνικό δίκτυο πολλαπλών επιπέδων με ένα κρυφό επίπεδο και τοπολογία 3-3-2.

Το νευρωνικό δίκτυο του σχήματος 1.6 λέγεται ότι είναι *πλήρως διασυνδεδεμένο* (fully connected), με την έννοια ότι κάθε κόμβος σε κάθε επίπεδο του δικτύου είναι συνδεδεμένος με κάθε κόμβο του επόμενου επιπέδου του δικτύου. Στην περίπτωση που κάποιες από τις συνδέσεις δεν υπάρχουν τότε λέμε ότι το δίκτυο είναι *μερικώς διασυνδεδεμένο* (partially connected).

1.3.3 Αναδραστικά Τεχνητά Νευρωνικά Δίκτυα (Recurrent Neural Networks (RNN))

Ένα Αναδραστικό Τεχνητό Νευρωνικό Δίκτυο (ΑΤΝΔ) διαφέρει από ένα Feedforward δίκτυο στο γεγονός ότι περιέχει έναν τουλάχιστον *βρόγχο ανατροφοδότησης*. Αυτό σημαίνει ότι σε έναν τουλάχιστον νευρώνα, το σήμα εξόδου του επηρεάζει το σήμα που έρχεται στην είσοδο του νευρώνα. Για παράδειγμα, ένα ΑΤΝΔ μπορεί να αποτελείται από ένα μόνο επίπεδο νευρώνων όπου κάθε νευρώνας επιστρέφει το σήμα εξόδου του σαν σήμα εισόδου σε όλους τους άλλους νευρώνες του επιπέδου (βλέπε σχήμα 1.7). Στο σχήμα αυτό δεν υπάρχουν κρυφοί νευρώνες ούτε βρόγχοι όπου ένας νευρώνας να επιστρέφει το σήμα εξόδου του σαν είσοδο στον εαυτό του. Ακόμα, στο σχήμα 1.8 παρουσιάζεται μια άλλη κλάση αναδρομικών νευρωνικών δικτύων που έχουν κρυφούς νευρώνες. Οι αναδρομικές συνδέσεις που φαίνονται στο σχήμα, ξεκινάνε τόσο από τους κρυφούς νευρώνες όσο και από τους νευρώνες εξόδου.

Η ύπαρξη αυτή των βρόγχων όπως φαίνονται στα σχήματα 1.7 και 1.8 επηρεάζουν σημαντικά τις επιδόσεις και την ικανότητα εκπαίδευσης του δικτύου. Επίσης, οι αναδρομικοί βρόγχοι περιλαμβάνουν και την χρήση στοιχείων καθυστέρησης (στο σχήμα συμβολίζονται με z^{-1}) έχοντας σαν αποτέλεσμα το δίκτυο να αποκτήσει μια δυναμική μη γραμμική συμπεριφορά.



Σχήμα 1.7: Αναδροστικό ΤΝΔ με ένα μόνο επίπεδο

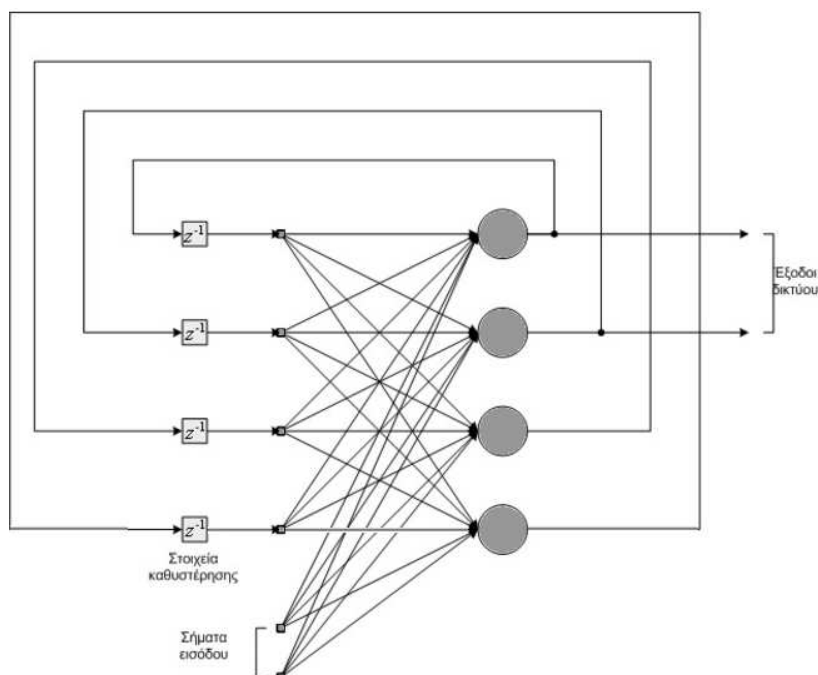
1.4 Διαδικασίες Μάθησης-Εκπαίδευσης

Η πιο σημαντική ιδιότητα των ΤΝΔ είναι η ικανότητά τους να *μαθαίνουν* από το περιβάλλον τους και να βελτιώνουν την απόδοσή τους μέσω της διαδικασίας μάθησης. Η βελτίωση αυτή γίνεται στην διάρκεια του χρόνου με κάποιο προκαθορισμένο μέτρο. Ένα ΤΝΔ μαθαίνει από το περιβάλλον του μέσω μιας επαναληπτικής διαδικασίας προσαρμογών των βαρών στις συνάψεις του δικτύου και στα κατώφλια του. Στην ιδανική περίπτωση, μετά από κάθε επανάληψη της διαδικασίας μάθησης το ΤΝΔ αποκτά περισσότερη γνώση για το περιβάλλον του.

Υπάρχουν πάρα πολλές δραστηριότητες που σχετίζονται με την έννοια της μάθησης έτσι είναι αρκετά δύσκολο να δοθεί ένας ακριβής ορισμός για την μάθηση [43]. Επιπλέον, ο ορισμός της μάθησης είναι θέμα οπτικής γωνίας και μπορεί να την δει κάποιος διαφορετικά από τους άλλους [50]. Στην συγκεκριμένη περίπτωση, ο ορισμός δίνεται από τους Mendel και McClaren [42].

Ορίζουμε την διαδικασία μάθησης στο πλαίσιο των τεχνητών νευρωνικών δικτύων ως:

Μάθηση είναι η διαδικασία με την οποία οι ελεύθερες παράμετροι ενός ΤΝΔ προσαρμόζονται μέσω μιας συνεχούς διαδικασίας λήψης ερεθισμάτων από το περιβάλλον στο οποίο είναι ενσωματωμένο το ΤΝΔ. Ο τύπος της



Σχήμα 1.8: Αναδραστικό ΤΝΔ με κρυφούς νευρώνες

μάθησης καθορίζεται από τον τρόπο με τον οποίο γίνονται οι αλληλαγές στις παραμέτρους.

Ο ορισμός αυτός της διαδικασίας μάθησης, συνεπάγεται την παρακάτω σειρά γεγονότων:

1. Το ΤΝΔ διεγείρεται από το περιβάλλον.
2. Το ΤΝΔ μεταβάλει τις τιμές των ελεύθερων παραμέτρων του ως αποτέλεσμα του ερεθισμάτων που δέχεται από το περιβάλλον.
3. Το ΤΝΔ αντιδρά με διαφορετικό τρόπο στο περιβάλλον, λόγω των αλλαγών που επήλθαν στην εσωτερική του δομή.

Για την εκπαίδευση των ΤΝΔ, δηλαδή για τον τρόπο με τον οποίο θα μεταβάλλονται οι ελεύθερες παράμετροι του, έτσι ώστε να επιτυγχάνεται η σωστή έξοδος για το ΤΝΔ ανάλογα με τα δεδομένα που εισέρχονται σε αυτό χρησιμοποιούνται οι *αλγόριθμοι εκπαίδευσης*.

Αλγόριθμος εκπαίδευση, καλείται ένα προκαθορισμένο σύνολο από κανόνες που χρησιμοποιούνται για την επίλυση ενός προβλήματος εκμάθησης.

Δεν υπάρχει ένας μοναδικός αλγόριθμος εκπαίδευσης κατάλληλος για όλες τις περιπτώσεις. Αντίθετα υπάρχει ένα σύνολο αλγορίθμων που χρησιμοποιούνται κατά περίπτωση ανάλογα με τα πλεονεκτήματα και τα μειονεκτήματά τους και το είδος του προβλήματος εκπαίδευσης. Το βασικό σημείο στο οποίο διαφέρουν οι αλγόριθμοι εκπαίδευσης είναι ο τρόπος με τον οποίο προσαρμόζουν τα βάρη w_{kj} των συνάψεων τους.

1.4.1 Μάθηση με διόρθωση σφάλματος (Error-Correction Learning)

Για να παρουσιάσουμε τον πρώτο κανόνα μάθησης, ας θεωρήσουμε έναν νευρώνα k , οποίος βρίσκεται στο επίπεδο εξόδου ενός ΤΝΔ, και έστω ότι βρισκόμαστε στην χρονική στιγμή n . $d_k(n)$ είναι η επιθυμητή απάντηση ή απάντηση στόχος (target response) για το νευρώνα k την χρονική στιγμή n . Έστω ότι η αντίστοιχη πραγματική απάντηση του νευρώνα k παριστάνεται με $y_k(n)$. Η απάντηση $y_k(n)$ προκαλείται από πρότυπο εισόδου, έστω $x(n)$, που εισάγεται στο επίπεδο εισόδου του ΤΝΔ, του οποίου στοιχείο είναι ο νευρώνας k . Το πρότυπο $x(n)$ και η επιθυμητή απάντηση $d_k(n)$ για το νευρώνα k αποτελούν ένα συγκεκριμένο παράδειγμα για το ΤΝΔ την χρονική στιγμή n . Γίνεται η υπόθεση ότι αυτό το παράδειγμα καθώς και όλα τα άλλα παραδείγματα που δίνονται στο δίκτυο, παράγονται από ένα πιθανοτικό περιβάλλον αλλά χωρίς να είναι γνωστή η κατανομή που έχουν τα στοιχεία του.

Στην πράξη, συνήθως, η πραγματική απάντηση $y_k(n)$ του νευρώνα k είναι διαφορετική από την απάντηση στόχο $d_k(n)$. Έτσι μπορούμε να ορίσουμε ένα σήμα σφάλματος ως την διαφορά μεταξύ της επιθυμητής απάντησης $d_k(n)$ και της πραγματικής απάντησης $y_k(n)$ του νευρώνα, όπως φαίνεται στην παρακάτω εξίσωση:

$$e_k(n) = d_k(n) - y_k(n).$$

Ο τελικός στόχος της εκπαίδευσης με διόρθωση σφάλματος είναι η ελαχιστοποίηση μιας *συνάρτησης κόστους* που βασίζεται στο σφάλμα $e_k(n)$, έτσι ώστε η πραγματική απάντηση κάθε νευρώνα στο δίκτυο να πλησιάζει την επιθυμητή απάντηση για το νευρώνα κατά μια στατιστική έννοια. Ένα κριτήριο που χρησιμοποιείται συχνά σαν συνάρτηση κόστους είναι το κριτήριο του μέσου τετραγωνικού σφάλματος, που ορίζεται σαν η μέση τετραγωνική τιμή του αθροίσματος του τετραγώνου των σφαλμάτων:

$$J = E \left[\frac{1}{2} \sum_k e_k^2(n) \right]$$

το οποίο είναι το άθροισμα των σφαλμάτων των νευρώνων εξόδου του ΤΝΔ. Η ελαχιστοποίηση της J ως προς τις παραμέτρους του δικτύου οδηγεί στην μέθοδο *gradient descent* (βλέπε [20] και [68]). Στην πράξη δεν είναι δυνατόν να βρεθεί η ακριβής λύση του προβλήματος βελτιστοποίησης με αυτή την μέθοδο και επομένως περιορίζομαστε στην εύρεση μιας προσεγγιστικής λύσης. Συγκεκριμένα, χρησιμοποιούμε την στιγμιαία τιμή του αθροίσματος των τετραγωνικών σφαλμάτων ως κριτήριο:

$$E(n) = \frac{1}{2} \sum_k e_k^2(n)$$

Σε αυτή την περίπτωση το δίκτυο βελτιστοποιείται ελαχιστοποιώντας την $E(n)$ ως προς τα βάρη των συνάψεων του δικτύου. Έτσι, σύμφωνα με τον κανόνα εκπαίδευσης με διόρθωση σφάλματος, η προσαρμογή $\Delta w_{kj}(n)$ που γίνεται στο βάρος w_{kj} την χρονική στιγμή n δίνεται από τον παρακάτω τύπο [67]:

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n)$$

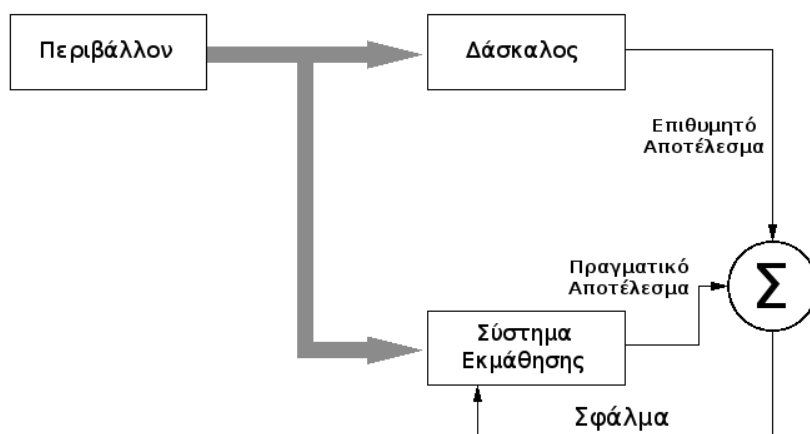
όπου η είναι μια θετική σταθερά που καθορίζει το *ρυθμό εκπαίδευσης*. Από την

εξίσωση αυτή βλέπουμε ότι η προσαρμογή που γίνεται στα βάρη είναι ανάλογη με το σφάλμα και το σήμα εισόδου της σύναψης που είναι το σήμα εξόδου του προσυναπτικού νευρώνα. Γενικά η εκπαίδευση με διόρθωση σφάλματος αποτελεί ένα σύστημα κλειστού βρόγχου. Ως εκ τούτου θα πρέπει η τιμή του ρυθμού εκπαίδευσης η να επιλέγεται έτσι ώστε να εξασφαλίζεται η σταθερότητα της διαδικασίας.

Η σημασία της τιμής του ρυθμού εκπαίδευσης είναι μεγάλη γιατί δεν επηρεάζει μόνο τον ρυθμό σύγκλισης της εκπαίδευσης αλλά και την ίδια την σύγκλιση της διαδικασίας. Αν το η είναι αρκετά μικρό η διαδικασία εκπαίδευσης προχωράει ομαλά αλλά είναι χρονοβόρα και το σύστημα αργεί να συγκλίνει σε μια σταθερή λύση. Από την άλλη πλευρά, αν το η επιλεγεί μεγάλο, ο ρυθμός της εκπαίδευσης θα επιταχύνεται αλλά τότε η διαδικασία εκπαίδευσης μπορεί να αποκλίνει και το σύστημα να γίνει ασταθές. Λόγω αυτού, έχουν προταθεί μέθοδοι που αλλάζουν κατάλληλα την τιμή του ρυθμού εκπαίδευσης έτσι ώστε η διαδικασία μάθησης να συγκλίνει στον επιθυμητό στόχο [37–40, 54, 65, 66]

1.4.2 Μάθηση με επίβλεψη (Supervised Learning)

Στην μάθηση με επίβλεψη είναι απαραίτητη η παρουσία ενός εξωτερικού, ως προς το δίκτυο, παράγοντα που μπορούμε να ονομάσουμε “δάσκαλο”. Στο σχήμα 1.9 παρουσιάζεται το πως επιδρά ο δάσκαλος στο δίκτυο και το περιβάλλον κατά την διαδικασία μάθησης. Ο δάσκαλος έχει την απαραίτητη γνώση για το περιβάλλον, που πρακτικά είναι ένα σύνολο από παραδείγματα εισόδου και την αντίστοιχη επιθυμητή έξοδο. Το ΤΝΔ δεν έχει καμία γνώση για το περιβάλλον. Αν υποθέσουμε ότι παρουσιάζουμε στον δάσκαλο και το δίκτυο ένα πρότυπο από το περιβάλλον, τότε λόγω της προηγούμενης γνώσης του δασκάλου για το περιβάλλον, θα είναι σε θέση να παρέχει στο δίκτυο την επιθυμητή απάντηση-έξοδο. Στη συνέχεια οι παράμετροι του δικτύου προσαρμόζονται ανάλογα με το πρότυπο που χρησιμοποιείται για την εκπαίδευση και το σφάλμα του δικτύου (δηλαδή την διαφορά μεταξύ της επιθυμητής εξόδου και της εξόδου που στην πράξη δίνει το δίκτυο). Η προσαρμογή αυτών των παραμέτρων, γίνεται επαναληπτικά, βήμα προς βήμα με στόχο το δίκτυο να μπορεί να προσομοιώσει τον δάσκαλο. Αν αυτό γίνει εφικτό, τότε μπορούμε να επιτρέψουμε στο δίκτυο να αλληλεπιδράσει με το περιβάλλον χωρίς την παρουσία του δασκάλου.



Σχήμα 1.9: Εκπαίδευση με Επίβλεψη.

Μια μορφή μάθησης με επίβλεψη είναι και η μάθηση με διόρθωση σφάλματος. Η μάθηση με επίβλεψη είναι ένα σύστημα κλειστού βρόγχου στο οποίο δεν περιλαμβάνεται το περιβάλλον μέσα στο οποίο λειτουργεί το δίκτυο. Σαν μέτρο απόδοσης για το σύστημα μπορούμε να θεωρήσουμε ένα είδος μέσου τετραγωνικού σφάλματος (π.χ. την μέση τιμή του αθροίσματος των τετραγωνικών σφαλμάτων) που ορίζεται σαν συνάρτηση των ελεύθερων παραμέτρων του συστήματος. Αυτή η συνάρτηση μπορεί να παρασταθεί σαν μια πολυδιάστατη *επιφάνεια σφάλματος* που σαν συντεταγμένες έχει τις ελεύθερες παραμέτρους του συστήματος. Κάθε λειτουργία του συστήματος υπό την επίβλεψη του δασκάλου αναπαριστάται σαν ένα σημείο στην επιφάνεια σφάλματος. Για να βελτιώνεται η απόδοση του συστήματος στη διάρκεια του χρόνου θα πρέπει το σημείο αυτό να έχει καθοδική πορεία προς κάποιο ελάχιστο, τοπικό ή ολικό, της επιφάνειας σφάλματος. Ένα σύστημα που μαθαίνει με επίβλεψη, έχει την δυνατότητα να μετακινεί το σημείο αυτό προς ένα ελάχιστο με την χρήση κάποιας πληροφορίας που έχει για την *κλίση* (gradient) της επιφάνειας σφάλματος που αντιστοιχεί στην συμπεριφορά του συστήματος την συγκεκριμένη χρονική στιγμή. Η κλίση της επιφάνειας σφάλματος σε οποιοδήποτε σημείο είναι ένα διάνυσμα που έχει κατεύθυνση προς την πιο *απότομη κάθοδο* (Steepest descent). Στην πράξη, όταν πρόκειται για μάθηση με επίβλεψη από παραδείγματα το σύστημα χρησιμοποιεί μια στιγμιαία προσέγγιση του διανύσματος κλίσης.

Ο ποιο γνωστός αλγόριθμος μάθησης με επίβλεψη είναι η Back-Propagation (BP) [21].

1.4.3 Μάθηση χωρίς επίβλεψη (Unsupervised Learning)

Στην περίπτωση της μάθησης χωρίς επίβλεψη ή αλλιώς μάθησης με αυτο-οργάνωση (Self-organization) δεν υπάρχει κάποιος εξωτερικός παράγοντας που επιβλέπει την διαδικασία μάθησης. Αυτό σημαίνει ότι δεν υπάρχουν παραδείγματα της συνάρτησης που πρέπει να μάθει το δίκτυο. Υπάρχει όμως ένα μέτρο, ανεξάρτητο από το εκάστοτε έργο που πρέπει να φέρει εις πέρας το ΤΝΔ, που μετράει την ποιότητα της αναπαράστασης που πρέπει να μάθει το δίκτυο. Οι ελεύθερες παράμετροι του δικτύου βελτιστοποιούνται ως προς αυτό το μέτρο. Όταν το δίκτυο “μάθει” τις στατιστικές ιδιότητες των προτύπων που του δίνονται σαν είσοδος, αναπτύσσει την ικανότητα να δημιουργεί εσωτερικές αναπαραστάσεις για την κωδικοποίηση των χαρακτηριστικών των προτύπων. Αποκτά δηλαδή την ικανότητα να δημιουργεί νέες κλάσεις αυτόματα [2].

Για την μάθηση χωρίς επίβλεψη μπορεί να χρησιμοποιηθεί ένας κανόνας ανταγωνιστικής μάθησης. Για παράδειγμα ένα ΤΝΔ με δύο επίπεδα, ένα επίπεδο εισόδου και ένα επίπεδο του οποίου οι νευρώνες θα ανταγωνίζονται για το ποιος θα ενεργοποιηθεί ανάλογα με τα χαρακτηριστικά του κάθε προτύπου που δίνεται σαν είσοδος στο ΤΝΔ.

1.5 Πολυεπίπεδα Εμπρόσθιας Τροφοδότησης ΤΝΔ ή Multilayer Perceptrons (MLP)

Ένα MLP δίκτυο αποτελείται από ένα σύνολο κόμβων εισόδου που αποτελούν το *επίπεδο εισόδου*, ένα ή περισσότερα *κρυφά επίπεδα* που αποτελούνται από νευρώνες που εκτελούν υπολογισμούς και ένα επίπεδο εξόδου που αποτελείται επίσης από

νευρώνες που εκτελούν υπολογισμούς. Το σήμα εισόδου (πρότυπο εισόδου) κινείται μέσα στο δίκτυο προς τα μπροστά, δηλαδή από το ένα επίπεδο στο επόμενο του.

Τα MLP συνήθως εκπαιδεύονται με κανόνες μάθησης με επίβλεψη. Ένας αλγόριθμος που χρησιμοποιείται πολύ συχνά για τον σκοπό αυτό είναι γνωστός ως αλγόριθμος Back-Propagation (BP) και βασίζεται στον κανόνα μάθησης με διόρθωση σφάλματος. Ο αλγόριθμος αυτός μπορεί να θεωρηθεί σαν γενίκευση του αλγορίθμου Ελάχιστων Μέσων Τετραγώνων (Least Mean Square Algorithm).

Ένα MLP έχει τρία σαφή χαρακτηριστικά :

1. Το μοντέλο κάθε νευρώνα στο δίκτυο περιλαμβάνει μία μη-γραμμική *συνάρτηση ενεργοποίησης* στην έξοδό του. Το σημαντικό σημείο το οποίο πρέπει να παρατηρηθεί είναι ότι η μη γραμμικότητα αυτή είναι ομαλή (smooth), δηλαδή παντού διαφορίσιμη. Μια κοινός χρησιμοποιούμενη μορφή μη γραμμικότητας που ικανοποιεί αυτή την απαίτηση είναι η *σιγμοειδής συνάρτηση* ή *λογιστική συνάρτηση* και ορίζεται ως :

$$y_j = \frac{1}{1 + \exp(-u_j)}$$

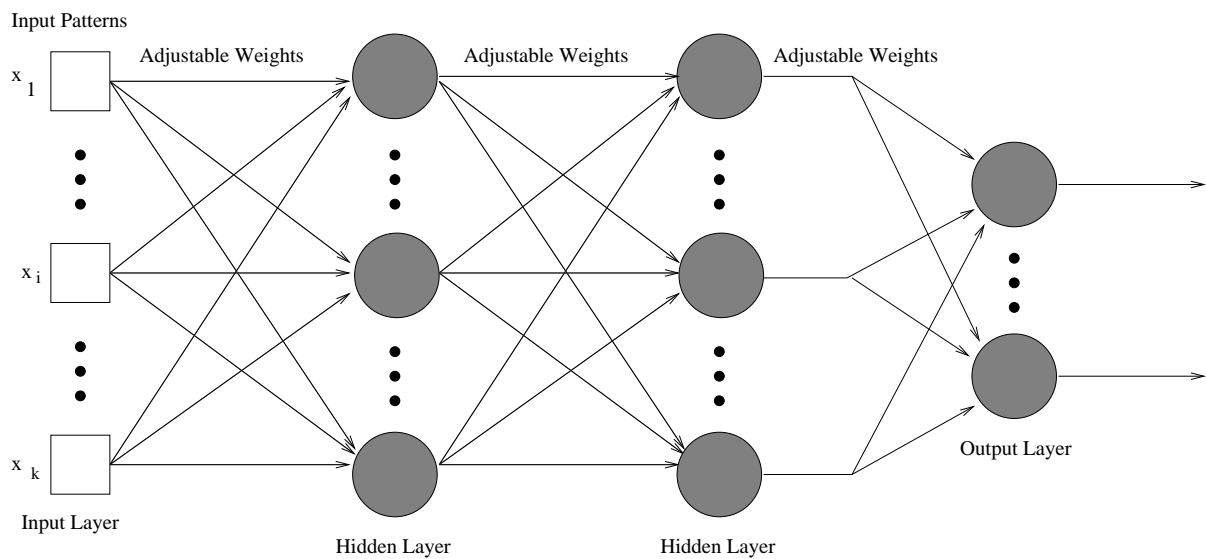
όπου u_j είναι η δραστηριότητα ενεργοποίησης του νευρώνα j και y_j είναι η έξοδος του νευρώνα. Η παρουσία της μη-γραμμικότητας είναι αρκετά σημαντική διότι σε διαφορετική περίπτωση η σχέση εισόδου-εξόδου του δικτύου θα μειωνόταν σε αυτή του ενός επιπέδου perceptron [21]. Επιπλέον, η χρήση της συνάρτησης αυτής έχει το πλεονέκτημα ότι προσομοιάζει την βιολογική φάση ανάσχεσης (refractory phase) στους πραγματικούς νευρώνες.

2. Το δίκτυο περιλαμβάνει ένα ή περισσότερα επίπεδα με *κρυφούς νευρώνες* που δεν αποτελούν μέρος είτε της εισόδου, είτε της εξόδου του δικτύου. Οι κρυφοί νευρώνες επιτρέπουν στο δίκτυο να “μαθαίνει” και να εκτελεί περίπλοκες εργασίες, εξάγοντας προοδευτικά τα χαρακτηριστικά εκείνα των προτύπων εισόδου που έχουν την μεγαλύτερη σημασία για την σωστή απάντηση του δικτύου.
3. Το δίκτυο παρουσιάζει υψηλό βαθμό συνδεσιμότητας που καθορίζεται από τις συνάψεις μεταξύ των νευρώνων του δικτύου. Μια αλλαγή στην συνδεσιμότητα του δικτύου απαιτεί μια αλλαγή στον πληθυσμό των συνδέσεων των συνάψεων ή στα βάρη τους.

Στο σχήμα 1.10 φαίνεται η αρχιτεκτονική ενός MLP δικτύου που έχει δύο κρυφά επίπεδα και είναι πλήρως διασυνδεδεμένο. Πλήρως διασυνδεδεμένο καλείται ένα δίκτυο του οποίου κάθε νευρώνας/κόμβος είναι συνδεδεμένος με όλους τους νευρώνες/κόμβους του προηγούμενου επιπέδου. Σε αυτής της μορφής τα δίκτυα το σήμα μεταδίδεται προοδευτικά από αριστερά προς τα δεξιά και από επίπεδο σε επίπεδο.

Στο δίκτυο μεταδίδονται δύο ειδών σήματα [51]:

1. *Σήματα συναρτήσεων* (Function signals). Ένα σήμα συνάρτησης είναι ένα σήμα εισόδου (ερέθισμα) που ξεκινάει από τους κόμβους εισόδου του δικτύου, διαδίδεται προς τα μπροστά, από νευρώνα σε νευρώνα, και καταλήγει στους νευρώνες εξόδου του δικτύου. Σε κάθε νευρώνα του δικτύου από όπου περνάει το σήμα υπολογίζεται σαν συνάρτηση όλων των εισερχόμενων σημάτων και των αντίστοιχων βαρών των συνάψεων που καταλήγουν στο συγκεκριμένο νευρώνα.



Σχήμα 1.10: Αρχιτεκτονική MLP δικτύου.

2. Σήματα σφάλματος (Error signals). Ένα σήμα σφάλματος, ξεκινάει από τους νευρώνες εξόδου του δικτύου και διαδίδεται προς τα πίσω από επίπεδο σε επίπεδο. Σε κάθε νευρώνα το σήμα αυτό υπολογίζεται από μια συνάρτηση που εξαρτάται από το σφάλμα.

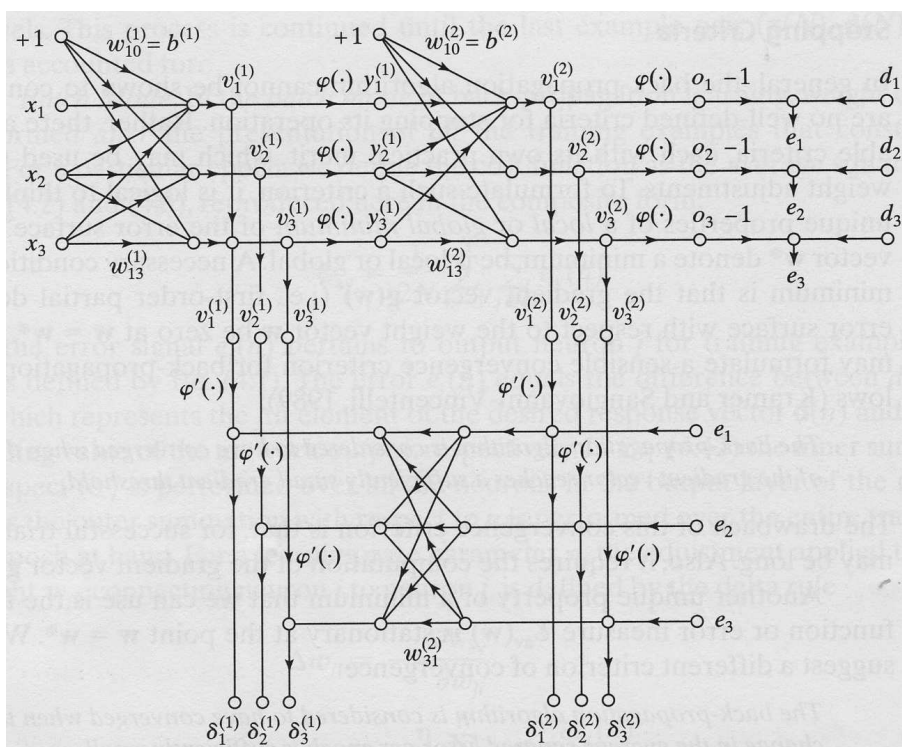
Οι νευρώνες που βρίσκονται στο επίπεδο εξόδου του ΤΝΔ, καλούνται νευρώνες εξόδου. Οι κόμβοι που βρίσκονται στο επίπεδο του ΤΝΔ από το οποίο ξεκινάει το σήμα ονομάζονται κόμβοι εισόδου. Όλοι οι άλλοι νευρώνες που βρίσκονται στα υπόλοιπα, ενδιάμεσα, επίπεδα του δικτύου λέγονται κρυφοί νευρώνες και τα επίπεδα κρυφά επίπεδα.

Κάθε κρυφός νευρώνας και κάθε νευρώνας εξόδου ενός MLP έχει σχεδιαστεί για να εκτελεί δύο υπολογισμούς:

1. Υπολογίζει το σήμα συνάρτησης που εμφανίζεται στην έξοδο του νευρώνα και εκφράζεται σαν μια συνεχής μη γραμμική συνάρτηση των σημάτων που εισέρχονται στο νευρώνα και των αντίστοιχων βαρών των συνάψεων που σχετίζονται με το νευρώνα.
2. Υπολογίζει μια στιγμιαία προσέγγιση του διανύσματος κλίσης (δηλαδή τις κλίσεις της επιφάνειας σφάλματος ως προς τα βάρη που σχετίζονται με τις συνάψεις που εισέρχονται στο νευρώνα), κατά την προς τα πίσω διάδοση του σήματος στο δίκτυο.

1.5.1 Ο αλγόριθμος Back-Propagation

Στο σχήμα 1.10 παρουσιάστηκε η αρχιτεκτονική ενός MLP. Η αντίστοιχη αρχιτεκτονική για την εκπαίδευση του δικτύου με τον αλγόριθμο Back Propagation παρουσιάζεται στο σχήμα 1.11 και περιλαμβάνει τόσο την προς τα εμπρός φάση της εκπαίδευσης όσο και την προς τα πίσω. Το MLP που φαίνεται στο πάνω κομμάτι του σχήματος αντιστοιχεί στην προς τα εμπρός φάση της διαδικασίας. Οι συμβολισμοί



Σχήμα 1.11: Η διαδικασία Back Propagation.

που χρησιμοποιούνται στο πάνω κομμάτι του σχήματος έχουν ως εξής:

- $\mathbf{w}^{(l)}$ = διάνυσμα βαρών των συνάψεων ενός νευρώνα στο επίπεδο l
- $\beta^{(l)}$ = κατώφλι ενός νευρώνα στο επίπεδο l
- $\mathbf{u}^{(l)}$ = διάνυσμα εσωτερικής δραστηριότητας των νευρώνων στο επίπεδο l
- $\mathbf{y}^{(l)}$ = διάνυσμα σημάτων συναρτήσεων των νευρώνων στο επίπεδο l

Ο δείκτης l κάθε επιπέδου ξεκινάει από το επίπεδο εισόδου ($l = 0$) μέχρι το δίκτυο εξόδου ($l = L$). Στο σχήμα 1.11 έχουμε $L = 3$ και καλούμε το L βάθος του δικτύου. Το κάτω κομμάτι του σχήματος 1.11 παρουσιάζει την προς τα πίσω φάση που αναφέρεται σαν *δίκτυο ευαισθησίας* και υπολογίζει τις τοπικές κλίσεις στον αλγόριθμο BP. Οι συμβολισμοί του κάτω τμήματος του σχήματος έχουν ως εξής:

- $\delta^{(l)}$ = διάνυσμα τοπικών κλίσεων των νευρώνων στο επίπεδο l
- \mathbf{e} = διάνυσμα σφάλματος που έχει ως στοιχεία τα e_1, e_2, \dots, e_q

Η διαδικασία που ακολουθεί ο αλγόριθμος Back-Propagation για να εκπαιδεύσει το δίκτυο έχει τα παρακάτω βήματα:

1. **Βήμα:** *Αρχικοποίηση.* Ανάθεσε σε όλα τα βάρη των συνάψεων και τα κατώφλια των νευρώνων του δικτύου μικρούς τυχαίους αριθμούς από την ομοιόμορφη κατανομή.
2. **Βήμα:** *Είσοδος των προτύπων για εκπαίδευση.* Δώσε στο δίκτυο για μια επανάληψη τα πρότυπα για την εκπαίδευση. Για κάθε πρότυπο του σύνολο εκτέλεσε

την παρακάτω αλληλουχία προς τα εμπρός και προς τα πίσω ενεργειών που περιγράφονται στο **Βήμα: 3** και **Βήμα: 4** αντίστοιχα.

3. **Βήμα:** *Προς τα εμπρός υπολογισμοί.* Έστω ότι ένα πρότυπο για εκπαίδευση συμβολίζεται με $[\mathbf{x}(n), \mathbf{d}(n)]$, όπου $\mathbf{x}(n)$ είναι το διάνυσμα του προτύπου που εισάγεται στο δίκτυο από το επίπεδο εισόδου και $\mathbf{d}(n)$ είναι η επιθυμητή απάντηση από το δίκτυο που δίνεται στους νευρώνες εξόδου του δικτύου. Υπολόγισε τις δυνατότητες ενεργοποίησης και τα σήματα των συναρτήσεων των νευρώνων προχωρώντας προς τα εμπρός στο δίκτυο από επίπεδο σε επίπεδο. Η εσωτερική δραστηριότητα $u_j^{(l)}(n)$ για τον νευρώνα j που βρίσκεται στο επίπεδο l του δικτύου δίνεται από

$$u_j^{(l)}(n) = \sum_{i=0}^m w_j^{(l)}(n) y_i^{(l-1)}(n)$$

όπου $y_i^{(l-1)}(n)$ είναι το σήμα συνάρτησης του νευρώνα i στο προηγούμενο επίπεδο $l - 1$ στην n -οστή επανάληψη και $w_j^{(l)}(n)$ είναι το βάρος της σύναψης που συνδέει τον νευρώνα j του επιπέδου l με το νευρώνα i του επιπέδου $l - 1$. Για $i = 0$ έχουμε $y_0^{(l-1)}(n) = -1$ και $w_0^{(l)}(n) = \beta_j^l(n)$, όπου $\beta_j^l(n)$ είναι η μεροληψία του νευρώνα j που βρίσκεται στο επίπεδο l . Έτσι με βάση τη σιγμοειδή συνάρτηση που χρησιμοποιήσαμε πιο πριν το σήμα συνάρτησης του νευρώνα j του επιπέδου l θα δίνεται από τον τύπο:

$$y_j^{(l)}(n) = \frac{1}{1 + \exp(-u_j^{(l)}(n))}$$

αν ο νευρώνας j βρίσκεται στο πρώτο κρυφό επίπεδο δηλαδή $l = 1$, θέσε

$$y_j^{(0)}(n) = x_j(n)$$

όπου $x_j(n)$ είναι το j -οστό στοιχείο του διανύσματος του προτύπου εισόδου (n). Ενώ, αν ο νευρώνας j βρίσκεται στο επίπεδο εξόδου, δηλαδή $l = L$, θέσε

$$y_j^{(L)}(n) = o_j(n)$$

όπου $o_j(n)$ είναι η έξοδος του δικτύου για το νευρώνα j . Στη συνέχεια υπολόγισε το σήμα σφάλματος

$$e_j(n) = d_j(n) - o_j(n)$$

Με $d_j(n)$ συμβολίζεται το j -οστό στοιχείο του διανύσματος της επιθυμητής απάντησης για το δίκτυο $\mathbf{d}(n)$.

4. **Βήμα:** *Προς τα πίσω υπολογισμοί.* Υπολόγισε τις τοπικές κλίσεις δ του δικτύου

προχωρώντας προς τα πίσω από επίπεδο σε επίπεδο ως εξής¹:

$$\delta_j^L(n) = e_j^L(n) o_j(n) [1 - o_j(n)]$$

για τον νευρώνα j στο επίπεδο εξόδου L

$$\delta_j^l(n) = y_j^{(l)}(n) [1 - y_j^{(l)}(n)] \sum_k \delta_k^{l+1}(n) w_{kj}^{l+1}(n)$$

για το νευρώνα j στο κρυφό επίπεδο L

Στη συνέχεια, προσάρμοσε τα βάρη των συνάψεων του δικτύου στο επίπεδο l σύμφωνα με τον γενικευμένο κανόνα του δ :

$$w_{ji}^l(n+1) = w_{ji}^l(n) + \alpha [w_{ji}^l(n) - w_{ji}^l(n-1)] + \eta \delta_j^l(n) y_i^{(l-1)}(n)$$

5. **Βήμα: Επανάληψη.** Επανάλαβε τους υπολογισμούς παρουσιάζοντας νέα πρότυπα εκπαίδευσης στο δίκτυο για κάθε επανάληψη μέχρις ότου οι ελεύθερες παράμετροι του δικτύου σταθεροποιηθούν και η συνάρτηση σφάλματος για όλο το σύνολο προτύπων εκπαίδευσης αποκτήσει την ελάχιστη τιμή της ή μια πολύ μικρή αποδεκτή τιμή. Η σειρά που θα τροφοδοτούνται τα πρότυπα στο δίκτυο σε κάθε εποχή θα πρέπει να μεταβάλλεται τυχαία.

1.6 Αναδραστικά Νευρωνικά δίκτυα

Τα αναδραστικά νευρωνικά δίκτυα χρησιμοποιούνται για την πρόβλεψη χρονοσειρών, δηλαδή για την αναγνώριση ακολουθιών όπου υπάρχει σειριακή συνοχή μεταξύ των στοιχείων. Αυτό συμβαίνει γιατί τα αναδραστικά νευρωνικά δίκτυα έχουν *δυναμική μνήμη* η οποία οφείλεται στα στρώματα ανατροφοδότησης που περιέχουν.

Ίσως η πρώτη πειραματική επίδειξη [21] για το αν τα αναδραστικά νευρωνικά δίκτυα μπορούν να μάθουν τα ενδεχόμενα που μπορούν να προκύψουν από πεπερασμένες γραμματικές έγινε το 1989 πάνω σε συμβολοσειρές. Πιο συγκεκριμένα το δίκτυο έπαιρνε σαν εισόδους συμβολοσειρές και καλούνταν να προβλέψει το επόμενο γράμμα σε ένα δεδομένο λεξιλόγιο συμβολοσειρών που είχε εκπαιδευτεί. Η δυσκολία σε αυτή τη διαδικασία είναι το στο λεξιλόγιο κάποιο γράμμα ή κάποια ακολουθία γραμμάτων μπορεί να ακολουθείται από διαφορετικά γράμματα. Το δίκτυο αποδείχτηκε ότι μπορεί να αναπτύξει κάποια αντίληψη για τη σειρά των γραμμάτων στις συμβολοσειρές πάνω στις οποίες είχε διδαχθεί.

Υπάρχουν διάφορες αρχιτεκτονικές αναδραστικών νευρωνικών δικτύων από τις οποίες θα περιγράψουμε την αρχιτεκτονική του *απλού αναδραστικού δικτύου* (simple recurrent network (SRNN)). Αυτό το μοντέλο είναι μία ειδική περίπτωση μίας κλάσης αναδραστικών νευρωνικών δικτύων που παρουσιάστηκε από τον Elman το 1990 [16].

Το απλό αναδραστικό νευρωνικό δίκτυο (ή δίκτυο Elman) έχει ένα κρυφό στρώμα νευρώνων του οποίου η έξοδος συνδέεται αναδραστικά με τις *μονάδες πλαισίου* (context units), ένα επιπλέον στρώμα νευρώνων το οποίο αποτελείται από μονάδες

¹Αν χρησιμοποιήσουμε σαν συνάρτηση ενεργοποίησης την $o(x) = \frac{1}{1+e^{-x}}$ τότε παρατηρούμε πως έχουμε για την παράγωγό της, $\frac{\partial o(x)}{\partial x} = \frac{1}{1+e^{-x}} = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}} \frac{e^{-x}}{1+e^{-x}} = \frac{1}{1+e^{-x}} \frac{1-1+e^{-x}}{1+e^{-x}} = \frac{1}{1+e^{-x}} \left[\frac{1}{1+e^{-x}} - \frac{1+e^{-x}}{1+e^{-x}} \right] = o(x)[o(x) - 1]$.

καθυστέρησης. Οι μονάδες πλαισίου αποθηκεύουν τις εξόδους των νευρώνων του κρυφού στρώματος για ένα βήμα υπολογισμών του αναδραστικού νευρωνικού δικτύου και στο επόμενο βήμα τις παρουσιάζει σαν επιπλέον εισόδους στο κρυφό στρώμα. Το κρυφό στρώμα τροφοδοτεί τέλος τους νευρώνες εξόδου. Οι νευρώνες που ανατροφοδοτούν το κρυφό στρώμα δεν είναι απλά μια πιστή καταγραφή της προηγούμενης κατάστασης των εσωτερικών νευρώνων καθώς συνεχίζουν να ανακυκλώνουν πληροφορίες μέσα στο δίκτυο με το πέρασμα του χρόνου.

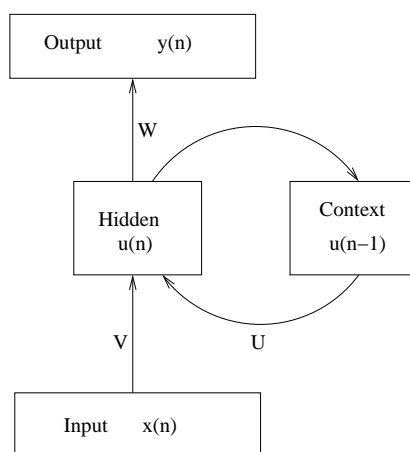
Ο Elman χρησιμοποίησε το απλό αναδραστικό δίκτυο για την εύρεση συνόρων για λέξεις σε ακολουθίες από συνεχόμενα φωνήματα χωρίς να υπάρχουν δομικοί περιορισμοί στην αναπαράστασή τους. Η είσοδος του νευρωνικού δικτύου αναπαριστούσε το τρέχων φώνημα και η έξοδος του δικτύου ήταν η πρόβλεψη για το ποιο θα είναι το επόμενο φώνημα. Οι μονάδες πλαισίου έδιναν στο δίκτυο τη δυνατότητα να έχει *δυναμική μνήμη* έτσι ώστε η πρόβλεψή του να λαμβάνει υπόψη τις πληροφορίες που δίνονταν από τη σειρά των φωνημάτων.

Περιγράφοντας σύντομα τη λειτουργία του απλού αναδραστικού δικτύου μπορούμε να πούμε πως αν $x(n)$ είναι η είσοδος του στο χρονικό βήμα n και $u(n)$ η έξοδος του κρυφού στρώματος, τότε η δυναμική συμπεριφορά του δικτύου φαίνεται από τις παρακάτω εξισώσεις:

$$u(n) = Vx(n) + Uu(n-1)$$

$$y(n) = Wu(n)$$

όπου $y(n)$ είναι η έξοδος του δικτύου στο n βήμα και f είναι η συνάρτηση που χαρακτηρίζει το κρυφό στρώμα.



Σχήμα 1.12: Αρχιτεκτονική SRNN δικτύου.

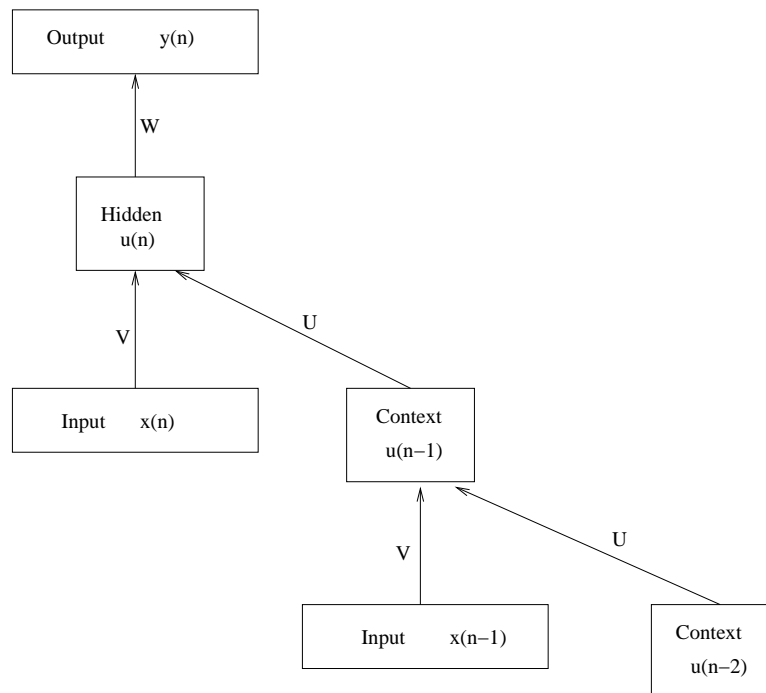
1.6.1 Ο αλγόριθμος Back-Propagation Through Time για τα δίκτυα Elman

Η μέθοδος εκπαίδευσης για τα αναδραστικά νευρωνικά δίκτυα είναι η μέθοδος Back-Propagation Through Time, (BPTT), η οποία έχει πολλές παραλλαγές, όλες

όμως έχουν σαν κεντρική ιδέα το *ξεδιπλώμα στο χρόνο* του αναδραστικού νευρωνικού δικτύου έτσι ώστε να είναι ισοδύναμο με ένα πολυεπίπεδο δίκτυο εμπρόσθιας τροφοδότησης. Στην παρούσα εργασία θα περιγράψουμε τη μέθοδο της *εποχικής εκπαίδευσης* (epochwise training).

Το δίκτυο Elman έχει ένα κρυφό στρώμα το οποίο τροφοδοτείται από τις μονάδες εισόδου $x(n)$ με ενδιάμεσα βάρη \mathbf{V} και έχει ως έξοδο το σήμα $u(n)$ το οποίο τροφοδοτεί στο στρώμα εξόδου με ενδιάμεσα βάρη \mathbf{W} . Το κρυφό στρώμα όμως τροφοδοτεί με την έξοδό του, $y(n)$, την είσοδο του αναδραστικού στρώματος, τις μονάδες πλαισίου, οι οποίες ανατροφοδοτούν το κρυφό στρώμα για τη λειτουργία του δικτύου στο χρονικό βήμα $n + 1$ μέσω των βαρών \mathbf{U} . Ενώ όμως βρισκόμαστε ακόμα στο βήμα n , το κρυφό στρώμα πρέπει να ανατροφοδοτηθεί από τις μονάδες πλαισίου με το σήμα εξόδου του κρυφού στρώματος στο βήμα $n - 1$, το $u(n - 1)$, μέσω των βαρών \mathbf{U} .

Στα δίκτυα Elman το στρώμα καθυστέρησης μπορεί να θεωρηθεί ως επιπλέον στρώμα εισόδου, έτσι ώστε το αναδραστικό τεχνητό νευρωνικό δίκτυο να μπορεί να θεωρηθεί σαν ένα δίκτυο εμπρόσθιας τροφοδότησης με στρώματα εισόδου που λαμβάνουν ως σήματα το $x(n)$ μέσω του διανύσματος βαρών \mathbf{V} και το $u(n - 1)$ μέσω του διανύσματος βαρών \mathbf{U} . Αναφορικά με την αρχική κατάσταση του δικτύου μπορούμε να θέσουμε $u(-1) = \vec{0}$.



Σχήμα 1.13: Ξεδιπλώμα στο χρόνο του παραπάνω απλού αναδραστικού δικτύου σε ένα ισοδύναμο εμπρόσθιας τροφοδότησης. Προσοχή θέλει το ότι τα βάρη \mathbf{V} και \mathbf{U} χρησιμοποιούνται μια φορά.

Με την παραπάνω διαδικασία έχουμε *ξεδιπλώσει στο χρόνο* το αναδραστικό δίκτυο σε ένα ισοδύναμο εμπρόσθιας τροφοδότησης. Μπορούμε να *ξεδιπλώσουμε* επιπλέον το δίκτυο έτσι ώστε να εκφράσουμε το $u(n - 1)$ σαν συνάρτηση του $x(n - 1)$ και του $u(n - 2)$ λαμβάνοντας έτσι ένα δίκτυο εμπρόσθιας τροφοδότησης το οποίο έχει σαν διανύσματα εισόδου τα $x(n)$, $x(n - 1)$ και $u(n - 2)$. Σε αυτό το σημείο

χρειάζεται προσοχή στο γεγονός ότι τα βάρη \mathbf{V} και \mathbf{U} παραμένουν τα ίδια καθώς ξεδιπλώνουμε το δίκτυο.

Το δίκτυο μπορεί να ξεδιπλωθεί στο χρόνο αυθαίρετα πολύ έτσι ώστε να έχει μνήμη που φτάνει μακρύτερα πίσω στον χρόνο. Στο ισοδύναμο δίκτυο εμπρόσθιας τροφοδότησης μπορούμε να εφαρμόσουμε τον αλγόριθμο Back Propagation για την εκπαίδευσή του, λαμβάνοντας όμως υπόψη πως τα βάρη \mathbf{V} και \mathbf{U} είναι τα ίδια στις ενώσεις των αντίστοιχων κόμβων για όσα επίπεδα και αν ξεδιπλώσουμε το δίκτυο.

Έτσι έχουμε τον αλγόριθμο Back Propagation Through Time σύμφωνα με τον οποίον οι αλλαγές που επιδέχονται τα βάρη \mathbf{V} και \mathbf{U} είναι το άθροισμα των αλλαγών που συμβαίνουν από τον αλγόριθμο Back Propagation στα τοπολογικά αντίστοιχα βάρη του ισοδύναμου δικτύου εμπρόσθιας τροφοδότησης. Συγκεκριμένα, αν θεωρήσουμε σαν T το πλήθος των επιπέδων που έχουμε ξεδιπλώσει το αναδραστικό τεχνητό νευρωνικό δίκτυο στο χρόνο, έχουμε ένα ισοδύναμο τεχνητό νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης με εισόδους τα $T - 1$ συνεχόμενα χρονικά βήματα της χρονοσειράς στην οποία θέλουμε να εκπαιδευτεί συν μία είσοδο για το περιεχόμενο των μονάδων πλαισίου του T χρονικού βήματος, δηλαδή τις παρακάτω εισόδους:

$$(\mathbf{x}(n), \mathbf{x}(n-1), \dots, \mathbf{x}((n-(T-1))), \mathbf{u}(n-T))$$

Έτσι εξηγείται το πώς το ΑΤΝΔ του σχήματος 1.12, το ξεδιπλώσαμε για $T = 2$ στο ισοδύναμο ΤΝΔ εμπρόσθιας τροφοδότησης του σχήματος 1.13, με εισόδους $\mathbf{x}(n), \mathbf{x}(n-1), \mathbf{u}(n-2)$.

Ο αλγόριθμος Back Propagation επιφέρει αλλαγές στα βάρη του ισοδύναμου ΤΝΔ εμπρόσθιας τροφοδότησης που αντιστοιχούν στα επίπεδα που έχει ξεδιπλωθεί το ΑΤΝΔ. Τις αλλαγές αυτές μπορούμε να τις συμβολίσουμε με $\delta\mathbf{V}^{(t)}$ και $\delta\mathbf{U}^{(t)}$ όπου $1 \leq t \leq T$ το επίπεδο ξεδιπλώματος καθώς και $\delta\mathbf{W}$ τα οποία δεν επηρεάζονται από τη διαδικασία του ξεδιπλώματος. Στο ΑΤΝΔ όμως, τα βάρη \mathbf{V} και \mathbf{U} είναι μοναδικά, έτσι οι τελικές αλλαγές που επιφέρει ο αλγόριθμος Back Propagation Through Time, $\delta\mathbf{V}$ και $\delta\mathbf{U}$, είναι, όπως είπαμε και παραπάνω, το άθροισμα των αλλαγών του Back Propagation στα αντίστοιχα επίπεδα, δηλαδή:

$$\delta\mathbf{V} = \sum_{t=1}^T \delta\mathbf{V}^{(t)}$$

$$\delta\mathbf{U} = \sum_{t=1}^T \delta\mathbf{U}^{(t)}$$

Η αντιστοίχιση του αλγορίθμου Back Propagation με τον αλγόριθμο Back Propagation Through Time ολοκληρώνεται αν θεωρήσουμε σαν εποχή εκπαίδευσης το σύνολο της χρονοσειράς στην οποία εκπαιδεύεται το δίκτυο. Το δίκτυο δηλαδή χρειάζεται τόσες εποχές για να εκπαιδευτεί όσες είναι και οι φορές που πρέπει να περάσει η χρονοσειρά για τη διόρθωση των βαρών του.

Το ξεδιπλωμα, όμως στον χρόνο του δικτύου από ένα σημείο και έπειτα μπορεί να επιφέρει αποτελέσματα που επηρεάζουν αρνητικά την ικανότητα του δικτύου να μαθαίνει. Αυτό το φαινόμενο ονομάζεται *φαινόμενο της εξαφανιζόμενης κλίσης* (vanishing gradient) [4] και συμβαίνει μιας και όλο και περισσότερες παράμετροι *συνεισφέρουν* στο άθροισμα της διόρθωσης των βαρών \mathbf{V} και \mathbf{U} με αποτέλεσμα να γίνεται δυσκολότερη η διαδικασία της εκπαίδευσης.

Σύνθεση μουσικής με πρόβλεψη με τη χρήση ΤΝΔ

2.1 Εισαγωγή

Στην αλγοριθμική σύνθεση μουσικής, μια απλή τεχνική περιλαμβάνει την επιλογή διαδοχικών νοτών σύμφωνα με έναν πίνακα μετάβασης που καθορίζει την πιθανότητα της επόμενης νότας σαν συνάρτηση του προηγούμενου περιεχομένου. Εδώ θα περιγράψουμε μία επέκταση αυτού του πίνακα μετάβασης, ένα αναδραστικό νευρωνικό δίκτυο που ονομάζεται CONCERT (**connectionist composer erudite tunes**). Το CONCERT έχει εκπαιδευτεί σε ένα σύνολο μουσικών κομματιών με σκοπό να εξάγει κανονικότητες που αφορούν είδη μουσικής και κατόπιν μπορεί να χρησιμοποιηθεί για να συνθέτει νέα μουσικά κομμάτια.

Το τμήμα αυτό της εργασίας βασίζεται στην εργασία του M. C. Mozer [47] που αποτελεί ίσως την ενδεδειγμένη έρευνα που έχει γίνει για το κατά πόσο τα τεχνητά νευρωνικά δίκτυα (χωρίς τη χρήση LSTM κυκλωμάτων [15, 19]) μπορούν να συνθέσουν μουσικά κομμάτια που να έχουν συνοχή και περιλαμβάνει την χρήση της πολυπλοκότερης μοντελοποίησης που έχει γίνει μέχρι σήμερα. Τα πειραματικά αποτελέσματα που περιγράφονται σε αυτή την εργασία έχουν ληφθεί από την προαναφερθείσα εργασία.

Για το σκοπό αυτό χρησιμοποιούνται αναπαραστάσεις της τονικότητας, της διάρκειας των νοτών και της αρμονικής δομής που βασίζονται στην ψυχολογία. Το CONCERT δοκιμάστηκε σε παραδείγματα που είχαν φτιαχτεί με απλούς κανόνες για να μάθει τις θεμελιώδεις δομές τους, ακόμη και εκεί που άλλες προσεγγίσεις απέτυχαν. Σε μεγαλύτερα πειράματα το CONCERT εκπαιδεύτηκε με κομμάτια του J.S. Bach και άλλα παραδοσιακά ευρωπαϊκά κομμάτια και μετά αφέθηκε να συνθέσει νέες μελωδίες. Παρόλο που οι συνθέσεις ήταν ευχάριστες και πιο αποδεκτές από συνθέσεις πινάκων τρίτης τάξης, οι συνθέσεις στερούνταν καθολικής συνοχής. Διάφορες τεχνικές δοκιμάστηκαν για να ξεπεραστεί το τελευταίο πρόβλημα.

Στη σύνθεση μουσικής, οι συνθέτες λαμβάνουν υπόψιν τους πολλές μουσικές συνθήκες μερικές από τις οποίες βασίζονται σε προσωπικές εμπειρίες, σε πολιτισμικές επιταγές και μερικές που μάλλον είναι συμπτωτικές. Όποιες και να είναι οι συνθήκες αυτές, δρουν για να περιορίσουν την μουσική δημιουργία, για παράδειγμα οι τονικότητες που αποτελούν την κλίμακα και οι συγχορδίες που είναι αποδεκτές πίσω από τις μελωδίες αυτές. Αν ελπίζουμε να δημιουργήσουμε ένα υπολογιστικό σύστημα που συνθέτει μουσική, πρέπει να λάβουμε υπόψιν μας τις παραπάνω συνθήκες. Είναι όμως δύσκολο να τυποποιήσουμε αυτή τη γνώση, αφού ακόμη και οι άνθρωποι που συνθέτουν δεν την αντιλαμβάνονται [35].

Η μουσική γνώση ενσωματώνεται στο CONCERT χρησιμοποιώντας δύο οδούς. Η πρώτη οδός είναι η εκπαίδευση του με μελωδίες από τις οποίες εξάγει συμπεράσματα για κανονικότητες που αφορούν την διαδοχή νοτών στις μουσικές φράσεις. Η δεύτερη οδός αφορά τον τρόπο αναπαράστασης της τονικότητας, της διάρκειας και της αρμονικής δομής που βασίζεται σε *ψυχοακουστικά* κριτήρια. Οι αναπαραστάσεις αυτές βασίζονται στη θεωρία γενικοποίησης που προτείνεται από τον Shepard [61] και εφοδιάζουν το CONCERT με *ψυχοακουστικούς* περιορισμούς με βάση τους οποίους κρίνει την ομοιότητα νοτών.

Τα πειράματα στα οποία θα αναφερθούμε γίνονται με μία φωνή που παράγει τη μελωδία και, όπου χρειαστεί, με αρμονική συνοδεία σε μορφή συγχορδιών. Οι μελωδίες αποτελούνται από 10 μέχρι περίπου 150 νότες. Μία πλήρης σύνθεση μουσικής πρέπει να περιγράφει την κάθε νότα με την τονικότητα και την διάρκεια, αλλά και να λαμβάνει υπόψιν πιο καθολικά κριτήρια όπως το τέμπο και οι δυναμικές. Στα πειράματα που θα διεξαχθούν εδώ περιγράφουμε μόνο την διαδοχή νοτών με τριάδες τονικότητας-διάρκειας-συγχορδίας.

Ξεκινάμε με μία εισαγωγική αναφορά στην αλγοριθμική σύνθεση μουσικής χρησιμοποιώντας Μαρκοβιανούς πίνακες μετάβασης. Θα δούμε τους περιορισμούς αυτής της προσέγγισης και πώς αυτοί μπορούν να ξεπεραστούν χρησιμοποιώντας το CONCERT.

2.2 Προσέγγιση του ζητήματος της αλγοριθμικής μουσικής με πίνακα μετάβασης

Μία απλή αλλά ενδιαφέρουσα τεχνική στην αλγοριθμική σύνθεση μουσικής είναι η επιλογή νοτών βάσει ενός *πίνακα μετάβασης* που καθορίζει την πιθανότητα για την επόμενη νότα σαν συνάρτηση της τρέχουσας νότας [14, 24, 34]. Για παράδειγμα, οι πιθανότητες μετάβασης που απεικονίζονται στον πίνακα 2.1 περιορίζουν την επιλογή της επόμενης νότας να είναι ένα βήμα πάνω ή κάτω την κλίμακα της C μείζονος. Παράγοντας μία σειρά νοτών με αυτή την κατανομή πιθανοτήτων θα έχουμε μία τυχαία μελωδική διαδρομή. Οι πίνακες μετάβασης μπορεί να είναι δημιουργημένοι με συγκεκριμένα κριτήρια, όπως στον πίνακα 2.1, ή να είναι ρυθμισμένοι για να ενσωματώνουν ένα συγκεκριμένο μουσικό είδος. Στην τελευταία περίπτωση τα στατιστικά στοιχεία συλλέγονται από ένα σύνολο από παραδείγματα (όπως θα δούμε το σύνολο εκπαίδευσης) και οι καταχωρήσεις των πινάκων μετάβασης καθορίζονται από τις πιθανότητες μετάβασης στα παραδείγματα αυτά.

Ο πίνακας μετάβασης αποτελεί μία στατιστική περιγραφή του συνόλου εκπαίδευσης και στις περισσότερες περιπτώσεις θα χάσει τις πληροφορίες από το σύνολο εκπαίδευσης. Αυτό θα γίνει πιο κατανοητό με το ακόλουθο παράδειγμα. Ας δώσουμε τις ακόλουθες δύο μελωδικές πορείες, $A B C$ και $E F G$. Ο πίνακας μετάβασης που κατασκευάζεται από τις προηγούμενες διαδοχές νοτών μας δείχνει πως το A θα πάει στο B με πιθανότητα 1, όπως και το B στο C κ.ο.κ. Στο συγκεκριμένο παράδειγμα ο πίνακας μετάβασης μπορεί να αναπαράγει το ακριβές σύνολο με το οποίο εκπαιδεύτηκε, αν του δοθεί η κατάλληλη αρχική νότα. Όμως αν υποθέσουμε τις μελωδικές πορείες $B A C$ και $D A E$ τότε ο πίνακας μετάβασης μπορεί να πει πως την A ακολουθεί η C ή η E με πιθανότητα $\frac{1}{2}$. Κατά συνέπεια, ο πίνακας δεν μπορεί να αναπαράγει ακριβώς το παράδειγμα.

Σε μελωδίες οποιασδήποτε πολυπλοκότητας, είναι σαφές πως η μουσική δομή

δεν μπορεί να περιγραφεί πλήρως από στατιστικές που εξετάζουν τμήματα της μελωδίας αυτής. Για να συλλέξουμε περισσότερες πληροφορίες για τη μουσική δομή, ο πίνακας μετάβασης μπορεί να γενικευτεί από ένα διδιάστατο, σε έναν n -διάστατο πίνακα. Σε αυτόν τον πίνακα η πιθανότητα για την εμφάνιση της επόμενης νότας εμφανίζεται ως πιθανότητα των $n - 1$ προηγούμενων νοτών. Γιαυτόν το λόγο, ο n -διάστατος πίνακας συχνά καλείται ως πίνακας $n-1$ τάξης. Όσο περισσότερο αυξάνουμε τον αριθμό των προηγούμενων νοτών που λαμβάνουμε υπόψιν, τόσο περισσότερο ο πίνακας γίνεται ευαίσθητος στο παρελθόν περιεχόμενό του, και έτσι μας δίνει μία πιο *πιστή αναπαράσταση*¹ του συνόλου εκπαίδευσης. Δυστυχώς η επέκταση του πίνακα μετάβασης καθεαυτόν τον τρόπο αντιμετωπίζει δύο προβλήματα. Πρώτον, το μέγεθος του πίνακα ανεβαίνει εκθετικά με το πλήθος των στοιχείων του να φτάνει εύκολα σε αφάνταστα μεγάλα μεγέθη. Για παράδειγμα, ας θεωρήσουμε πως έχουμε επιλέξει να διαχειριστούμε 50 διαφορετικές τονικότητες, 10 διάρκειες νοτών και έναν πίνακα τάξης 3, δηλαδή πίνακα τεσσάρων διαστάσεων. Όλες οι παραπάνω επιλογές είναι πολύ μειοπαθείς και όμως για να γίνει μία επιλογή από αυτόν τον πίνακα χρειάζονται 7.5 δισεκατομμύρια καταχωρίσεις. Δεύτερον, ένας πίνακας υψηλής τάξης επισκιάζει τις δομές που μπορούν να αναπαρασταθούν από πίνακες πρώτης τάξης, οι οποίες είναι πολλές σε ένα μουσικό κομμάτι. Ας επεξεργαστούμε το ακόλουθο το παράδειγμα με την ακόλουθη μελωδία

$A F G B F G C F G D F G\# E F G$

	C	D	E	F	G	A	B
C	0	.5	0	0	0	0	.5
D	.5	0	.5	0	0	0	0
E	0	.5	0	.5	0	0	0
F	0	0	.5	0	.5	0	0
G	0	0	0	.5	0	.5	0
A	0	0	0	0	.5	0	.5
B	.5	0	0	0	0	.5	0

Πίνακας 2.1: Πίνακας πιθανότητας μετάβασης από τη προσωρινή νότα, που βρίσκεται στη πρώτη γραμμή, στην επόμενη που βρίσκεται στη πρώτη στήλη.

Για να αναπαραστήσουμε την παραπάνω μελωδία θα μπορούσαμε να χρησιμοποιήσουμε έναν πίνακα τάξης τρία έτσι ώστε, για παράδειγμα, η ακολουθία νοτών $G B F$ θα ξέρουμε πως έχει σαν συνέχεια την G . Παρόλα αυτά, υπάρχουν και κανονικότητες πρώτης τάξης στην παραπάνω ακολουθία που δεν μπορούν να συναχθούν από δομές τρίτης τάξης, όπως, για παράδειγμα, η F σχεδόν πάντα ακολουθείται από G . Έτσι, ένας πίνακας μετάβασης τρίτης τάξης θα είχε δυσκολία να αποφανθεί για την συνέχεια της ακολουθίας $A A F$, ενώ ένας πίνακας τάξης ένα θα απαντούσε G με σχετική βεβαιότητα. Υπάρχει λοιπόν, ένα αντίτιμο για την ικανότητα της πιστής

¹Χρησιμοποιώντας την ορολογία του Smolensky, με τον όρο πιστή αναπαράσταση εννοούμε ότι τα αντικείμενα προς αναπαράσταση μπορούν να αναπαραχθούν με ακρίβεια από την αναπαράσταση. Ένας πιστός πίνακας μετάβασης, σύμφωνα με αυτή την ορολογία, είναι ένας πίνακας ο οποίος χρησιμοποιώντας κάποιες αρχικές νότες ενός παραδείγματος καταφέρνει με βεβαιότητα να βρίσκει την επόμενη.

αναπαράστασης του συνόλου εκπαίδευσης που μας δίνει ένας υψηλής τάξης πίνακας εναντίον της ικανότητας για γενίκευση και ανταπόκριση σε καινούριο περιεχόμενο που μας δίνει ένας πίνακας χαμηλότερης τάξης [31].

Ο Kohonen [26, 27] έχει προτείνει ακριβώς αυτό το σχέδιο. Αυτό είναι ένας συμβολικός αλγόριθμος ο οποίος, δοθέντος ενός συνόλου εκπαίδευσης, παράγει ένα σύνολο κανόνων - μία γραμματική ευαίσθητη στο προηγούμενο περιεχόμενο - επαρκές για να αναπαράγει τις περισσότερες ή όλες τις δομές που κληρονομήθηκαν από το σύνολο αυτό. Αυτοί οι κανόνες είναι της μορφής *προηγούμενο_περιεχόμενο* → *επόμενη_νότα*, όπου το *προηγούμενο_περιεχόμενο* είναι μία συμβολοσειρά που αποτελείται από μία ή περισσότερες προηγούμενες νότες και *επόμενη_νότα* είναι η νότα που συνεπάγεται του περιεχομένου αυτού. Μιας και το μήκος του προηγούμενου περιεχομένου ποικίλει, ανάλογα με τον κανόνα που θα επιλεγεί, ο αλγόριθμος αυτός επιτρέπει διάφορες ποσότητες γενίκευσης και ακρίβειας. Ο αλγόριθμος επιχειρεί να δημιουργήσει ντετερμινιστικούς κανόνες - κανόνες που πάντα εφαρμόζονται στο δεδομένο προηγούμενο περιεχόμενο. Δηλαδή, ο αλγόριθμος δε θα ανακαλύψει την κανονικότητα $F \rightarrow G$ στο παραπάνω παράδειγμα μιας και δεν είναι απόλυτος. Κάποιος θα μπορούσε να σκεφτεί μία επέκταση του αλγορίθμου έτσι ώστε να παράγει απλούς κανόνες, όπως $F \rightarrow G$, μαζί με εξαιρέσεις, όπως $D F \rightarrow G\#$, αλλά η συμβολική φύση του αλγορίθμου δεν τον αφήνει να χρησιμοποιεί στατιστικές πληροφορίες των δεδομένων. Τέτοιες πληροφορίες δεν είναι απαραίτητες αν ο σκοπός του αλγορίθμου είναι να αναπαράγει επακριβώς το σύνολο εκπαίδευσης. Αν όμως δούμε την σύνθεση μουσικής σαν μία καθαρά τυχαία διαδικασία, δεν πρέπει να μοντελοποιούμε όλες τις λεπτομέρειες του συνόλου εκπαίδευσης. Αντί γιαυτό, ο στόχος θα έπρεπε να είναι να κρατούμε τις πιο σημαντικές - τις στατιστικές παραδείγματος χάριν - δομικές ιδιότητες του συνόλου εκπαίδευσης.

Και οι δύο προσεγγίσεις που είδαμε μέχρι τώρα, οι πίνακες μετάβασης και η μουσική γραμματική του Kohonen, πάσχουν από δύο επιπλέον μειονεκτήματα. Πρώτον, και οι δύο αλγόριθμοι είναι έτσι σχεδιασμένοι ώστε να μη μπορεί να χρησιμοποιηθεί η νότα n για να προβλέψει την νότα $n + i$, αν δεν ληφθούν υπόψιν και όλες οι υπόλοιπες παρεμβαλλόμενες νότες, $n + 1 \dots n + i - 1$. Κάποιος θα περίμενε πως το περιεχόμενο που μπορεί να προβλέψει καλύτερα μια νότα είναι η αμέσως προηγούμενη της, υπάρχουν όμως και περιπτώσεις που οι νότες $n \dots n + k$ αποτελούν καλύτερη πρόβλεψη της νότας $n + i$ από τις $n + k + 1 \dots n + i - 1$. Τέτοια παραδείγματα αποτελούν οι μελωδικές φράσεις που εναλλάσσονται μεταξύ ψηλότερων και χαμηλότερων τονικοτήτων, όπως σε παρτίτες σόλο βιολιού του J.S.Bach. Το δεύτερο, και μάλλον πιο σοβαρό μειονέκτημα, είναι πως αυτή η συμβολική αναπαράσταση νοτών δεν διευκολύνει γενικεύσεις από ένα μουσικό περιεχόμενο σε ένα ομοίως αντιληπτό. Για παράδειγμα, η μουσική ομοιότητα των οκτάβων και η αφηρημένη έννοια της 'μικρής τρίτης' δεν κωδικοποιούνται με την παραπάνω αναπαράσταση.

Οι αλγόριθμοι που περιλαμβάνουν εκπαίδευση νευρωνικών δικτύων προσφέρουν την προοπτική να ξεπεραστούν διάφοροι περιορισμοί που προκύπτουν από τους πίνακες μετάβασης και την μουσική γραμματική του Kohonen. Αν και μπορούν να χαρακτηριστούν ως επέκταση των πινάκων μετάβασης, οι αλγόριθμοι που προκύπτουν από νευρωνικά δίκτυα μπορούν να ανακαλύπτουν συσχετισμένες δομές και στατιστικές κανονικότητες σε χρονοσειρές σειρές [16, 45]. Ακριβώς όπως οι πίνακες μετάβασης, η προσέγγιση με νευρωνικά δίκτυα χρησιμοποιεί έναν πίνακα για να υπολογίσει την πιθανότητα της εμφάνισης της επόμενης νότας, όμως λειτουργεί με πιο ευέλικτες αρχές: η μορφή της συνάρτησης μετάβασης μπορεί να επιτρέψει μετα-

βλητό μήκος προηγούμενου περιεχομένου που θα καθορίσει την επόμενη νότα, τον καθορισμό της επόμενης νότας όχι απαραίτητα από γειτονικό περιεχόμενο και τον συνδυασμό κανονικοτήτων υψηλής και χαμηλής τάξης.

Η προσέγγιση των νευρωνικών δικτύων προσφέρει επίσης καλύτερη γενίκευση μέσω της κατανεμημένης αναπαράστασης [57]. Σε μια τοπική αναπαράσταση, που η κάθε νότα περιγράφεται από το σύμβολό της, οι στατιστικές κανονικότητες που συμπεραίνονται στην πραγματικότητα δεν έχουν να κάνουν με νότες. Από την άλλη, με την αναπαράσταση κατανομής, που κάθε νότα περιγράφεται από ένα σύνολο στοιχείων με συνεχείς τιμές, το είδος των συνοχών που συμπεραίνονται στην πραγματικότητα δεν έχουν να κάνουν τα στοιχεία αυτά, αλλά με τις δομές που τα συνδέουν μεταξύ τους.

2.3 Σύνθεση νότας προς νότα

Το CONCERT, το οποίο θα μελετήσουμε, χρησιμοποιεί τη σύνθεση νότας προς νότα όπως και οι πίνακες μετάβασης του Markon και η μουσική γραμματική του Kohonen. Προσεγγίσεις του ζητήματος της σύνθεσης μουσικής νότας προς νότα με τη χρήση νευρωνικών δικτύων έχουν ξαναγίνει στο παρελθόν [9, 62, 63], το CONCERT όμως χρησιμοποιεί πιο εκλεπτυσμένες τεχνικές βελτιστοποίησης και εκπαίδευσης. Σκοπός αυτής της έρευνας είναι να δούμε πόσο μακριά μπορεί να φτάσει η σύνθεση νότας προς νότα.

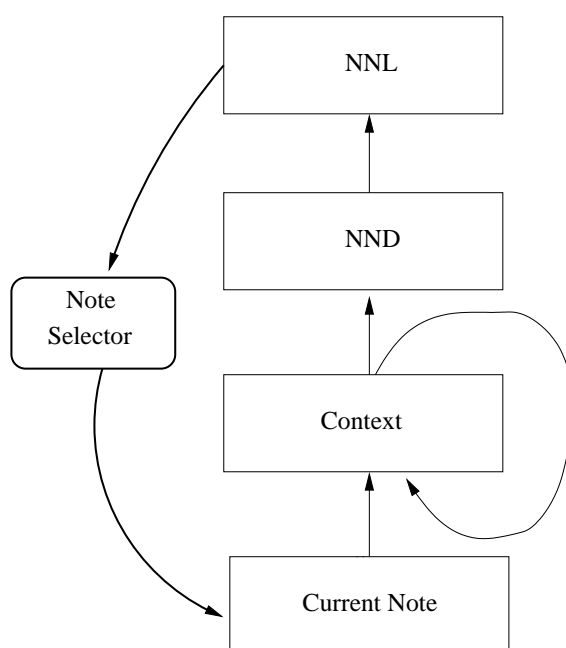
Παρά την αυξημένη έρευνα στη σύνθεση μουσικής νότας προς νότα, αυτή η τεχνική δείχνει πως δεν μπορεί να πετύχει ολοκληρωτικά. Η μουσική έχει μία πλούσια και ιεραρχική δομή που, από επίπεδα νοτών σε ένα θέμα, από θέματα μέσα σε φράσεις, από φράσεις μέσα σε κινήσεις, από κινήσεις μέσα σε μια συμφωνία. Κάποιος θα μπορούσε να διαφωνήσει πως η γραμμική σύνθεση θα μπορούσε να ελέγχει όλες αυτές τις δομές. Τα νευρωνικά δίκτυα όμως, δείχνουν να τα καταφέρνουν και με την συγκεκριμένη έρευνα θα δούμε στην πράξη τον βαθμό στον οποίον το καταφέρνουν.

Αυτή η γραμμική τεχνική έχει να επιδείξει εντυπωσιακά αποτελέσματα για την επεξεργασία της φυσικής γλώσσας [16, 17]. Νευρωνικά δίκτυα έχουν εκπαιδευτεί σε συμβολοσειρές ή λέξεις αποτελούμενες από γράμματα και κατάφεραν να βγάλουν συμπεράσματα για γραμματική ή σημασιολογική δομή. Αυτή η εργασία όμως, περιορίστηκε σε επίπεδο προτάσεων. Είναι δύσκολο να φανταστούμε τον τρόπο με τον οποίο θα μπορούσε ένα νευρωνικό δίκτυο να διαχειριστεί σημασιολογική δομή σε παραγράφους ή ακόμα και σε μικρές ιστορίες. Ευτυχώς η μουσική δεν είναι σαν στην φυσική γλώσσα σε θεμελιώδη επίπεδα: το πλήθος των ατομικών στοιχείων είναι πεπερασμένο, σχετικά μικρό και σαφές και οι *ψυχοακουστικές* και μορφολογικές κανονικότητες αφθονούν. Αυτό έχει σαν αποτέλεσμα οι περιορισμοί μεταξύ των στοιχείων να είναι πιο ισχυροί. Έτσι φαίνεται πως, τουλάχιστον σε κάποιο επίπεδο, η διαχείριση της μουσικής γλώσσας φαίνεται πραγματοποιήσιμη.

2.4 Αρχιτεκτονική του CONCERT

Το CONCERT είναι ένα νευρωνικό δίκτυο το οποίο, ακολουθώντας την διαδικασία εκπαίδευσης που περιγράφεται παρακάτω, προσαρμόζει τα βάρη του έτσι ώστε να μπορεί να εντοπίζει την επόμενη μιας νότας για ένα σύνολο από παραδείγματα

εκπαίδευσης. Κάθε σύνολο εκπαίδευσης αποτελείται από μία σειρά νοτών. Η προσωρινή νότα παρουσιάζεται στο στρώμα εισόδου και η επόμενη νότα που προβλέπει παρουσιάζεται στο στρώμα εξόδου. Οι πληροφορίες που το CONCERT διαχειρίζεται για την κάθε νότα είναι η *τονικότητα*, η *διάρκεια* και η *αρμονική συνοδεία* της νότας αυτής. Όπως φαίνεται στο σχήμα 2.1, η επόμενη νότα κωδικοποιείται με δύο τρόπους: το στρώμα κατανομής της επόμενης νότας (*next-note-distributed*) ή NND, το οποίο περιέχει τρεις μονάδες που περιέχουν πιθανοτικές αναπαραστάσεις κατανομής της τονικότητας, της διάρκειας και της αρμονίας, και το στρώμα τύπου της επόμενης νότας *next-note-local* ή NNL, το οποίο περιέχει μία ξεχωριστή μονάδα για κάθε τονικότητα, διάρκεια και αρμονία. Η αναπαράσταση της νότας στο στρώμα NND όσο και το στρώμα εισόδου βασίζεται σε ψυχοφυσικές μελέτες που θα περιγράψουμε παρακάτω.



Σχήμα 2.1: Η αρχιτεκτονική του CONCERT

Το στρώμα του περιεχομένου μπορεί να αναπαραστήσει κατάλληλες οπτικές της ιστορίας της εισόδου, δηλαδή του περιεχομένου της δεδομένης χρονικής στιγμής την οποία θα γίνει η πρόβλεψη. Μόλις μία νέα νότα παρουσιάζεται στο στρώμα εισόδου, το προσωρινό *σχέδιο δραστηριότητας* που υπάρχει στο στρώμα πλαισίου (βλ. κεφάλαιο 1.6), συμπληρώνεται από τη νότα αυτή και σχηματίζει μια νέα αναπαράσταση πλαισίου. Δηλαδή

$$\vec{c}(n) = f(\vec{c}(n-1), \vec{x}(n))$$

όπου το $\vec{x}(n)$ είναι το διάνυσμα που αναπαριστά την n -οστή νότα στο στρώμα εισόδου, $\vec{c}(n)$ είναι το περιεχόμενο της δραστηριότητας πλαισίου στο βήμα n και f είναι ένα μέλος τη κλάσης συναρτήσεων που τίθενται σε εφαρμογή από το νευρωνικό δίκτυο στο στρώμα πλαισίου. Στην αρχή της κάθε διαδικασίας ισχύει $\vec{c}(-1) = \vec{0}$.

Το CONCERT μπορεί να συμπεριφερθεί σαν ένας πίνακας k -οστής τάξης, με τη συνάρτηση f ορισμένη ώστε να δρα πάνω σε μια στοιβά k στοιχείων στο στρώμα πλαισίου. Έτσι, οι συνδέσεις του στρώματος περιεχομένου με το στρώμα εξόδου

θα έπρεπε να είναι ρυθμισμένες ώστε κάθε συνδυασμός νοτών στο προηγούμενο περιεχόμενο να έδινε μια πιθανότητα για την επόμενη νότα, μέσω ενός πίνακα αναζήτησης. Η αρχιτεκτονική του CONCERT όμως, επιτρέπει πιο γενική προσέγγιση απότι οι πίνακες μετάβασης, μιας και η συνάρτηση f δεν περιορίζεται στο να δρα σε k στοιχεία και η απεικόνιση από το στρώμα πλαισίου στο στρώμα εξόδου δεν είναι ένας αυθαίρετος πίνακας αναζήτησης. Η διαδικασία εκπαίδευσης προσπαθεί να βρει το κατάλληλο σύνολο συνδέσεων που να είναι επαρκές για την πρόβλεψη της επόμενης νότας, μέσα από πολλούς δυνατούς συνδυασμούς. Η διαδικασία αυτή περιλαμβάνει τον καθορισμό του ποιες από τις παραμέτρους του στρώματος εισόδου είναι υπεύθυνες για μελλοντικές προβλέψεις και την κατάλληλη κατασκευή της συνάρτησης f .

Μόλις το CONCERT εκπαιδευτεί, μπορεί να λειτουργήσει σε *λειτουργία σύνδεσης*, δημιουργώντας νέα κομμάτια. Αυτό γίνεται με την τροφοδότηση του CONCERT με λίγες αρχικές νότες, μάλλον με τις αρχικές νότες ενός κομματιού από το σύνολο εκπαίδευσης. Από εδώ και πέρα, το CONCERT μπορεί να ανατροφοδοτεί το στρώμα εισόδου του, με τη νότα που βγαίνει από το στρώμα εξόδου του, συνεχίζοντας έτσι να κάνει προβλέψεις και να παράγει νότες, συνεχίζοντας το κομμάτι. Η έξοδος του CONCERT δεν είναι μια συγκεκριμένη νότα, είναι μία κατανομή πιθανοτήτων για την επιλογή μεταξύ διαφόρων υποψήφιων νοτών. Είναι, λοιπόν, απαραίτητο να υπάρχει μία διαδικασία επιλογής μίας νότας συναρτήσει της κατανομής αυτής. Αυτό απεικονίζεται στο σχήμα 2.1.

2.4.1 Κανόνες ενεργοποίησης μονάδας

Ο κανόνας ενεργοποίησης της μονάδας πλαισίου είναι

$$c_i(n) = s \left[\sum_j w_{ij} x_j(n) + \sum_j v_{ij} c_j(n-1) \right]$$

όπου $c_i(n)$ είναι η δραστηριότητα της μονάδας πλαισίου i στο βήμα n , $x_j(n)$ είναι η δραστηριότητα της μονάδας εισόδου j στο βήμα n , w_{ij} είναι το βάρος της σύνδεσης από τη μονάδα εισόδου j στη μονάδα περιεχομένου i , v_{ij} είναι το βάρος σύνδεσης της μονάδας i στη μονάδα j μέσα στη μονάδα περιεχομένου, και s είναι η λογιστική συνάρτηση ενεργοποίησης

$$s(t) = \frac{1}{1 + e^{-t}}$$

με κλίμακα $(-1, 1)$. Οι μονάδες στο στρώμα NND ακολουθούν παρόμοιο κανόνα:

$$nnd_i(n) = s \left[\sum_j u_{ij} c_j(n) \right]$$

όπου nnd_i είναι η δραστηριότητα της i -οστής μονάδας NND στο βήμα n και u_{ij} είναι το βάρος της σύνδεσης από τη μονάδα πλαισίου j στη μονάδα NND i .

Οι αναπαραστάσεις στα στρώματα NND και NNL μπορούν να σπάσουν σε ένα διάγραμμα τριών στοιχείων που αντιστοιχούν στην τονικότητα, τη διάρκεια και την συνοδευτική συγχορδία. Τώρα θα περιγράψουμε την μετατροπή της τονικότητας από το NND στρώμα στο NNL. Η μετατροπή της διάρκειας και της τονικότητας είναι παρόμοια. Η μετατροπή της τονικότητας επιτυγχάνεται αν πρώτα υπολογίσουμε

την απόσταση της τονικότητας που παρουσιάζεται στο στρώμα NND, δηλαδή του διανύσματος $nndp(n)$, και του διανύσματος στόχου της κάθε τονικότητας i , \vec{p}_i :

$$d_i = \|nndp(n) - \vec{p}_i\|$$

όπου το $\|\cdot\|$ συμβολίζει την στάθμη L_2 . Όσο μικρότερη η τιμή αυτής της απόστασης τόσο περισσότερο η αναπαράσταση του στρώματος NND ταιριάζει με τη νότα που έχει τεθεί σαν στόχος. Η ενεργοποίηση της μονάδας NNL που αντιστοιχεί στη νότα i , δηλαδή η nnp_i , αυξάνεται όσο η απόσταση μειώνεται, και συγκεκριμένα:

$$nnp_i(n) = \frac{e^{-d_i}}{\sum_j e^{-d_j}}$$

Αυτή η εκθετική μετατροπή παράγει μία μορφή δράσης στις μονάδες του στρώματος NNL σύμφωνα με την οποία κάθε μονάδα έχει δραστηριότητα με εύρος $(0, 1)$ και η δραστηριότητα όλων των μονάδων αθροίζει στο 1 [10]. Έτσι, η μορφή δράσης του στρώματος NNL μπορεί να ερμηνευτεί ως μία συνάρτηση κατανομής η οποία, εν προκειμένω, μας δείχνει την πιθανότητα η επόμενη νότα να έχει συγκεκριμένη τονικότητα. Μία παρόμοια μετατροπή συμβαίνει στο στρώμα NNL για να αναπαρασταθεί η διάρκεια της νότας και η συγχορδία που την συνοδεύει. Το μέτρο της απόστασης και η εκθετική συνάρτηση βασίζονται σε ψυχολογικές θεωρίες [61] και θα ασχοληθούμε με αυτές προσεχώς.

2.4.2 Διαδικασία εκπαίδευσης

Το CONCERT εκπαιδεύεται με τη χρήση της μεθόδου προς τα πίσω ολισθήσεως του σφάλματος (Back Propagation) [57] με την οποία ρυθμίζονται τα βάρη των συνδέσεων εντός του, έτσι ώστε το δίκτυο να μπορεί να κάνει προβλέψεις νοτών για τα παραδείγματα του συνόλου εκπαίδευσης. Ο αλγόριθμος για τη μέθοδο εκπαίδευσης απαιτεί πρώτα τον καθορισμό του μέτρου της αποδοτικότητας του δικτύου, δηλαδή το πόσο σωστά προβλέπει το δίκτυο την επόμενη νότα στα παραδείγματα εκπαίδευσης. Όπως συνήθως, και εδώ χρησιμοποιείται η τετραγωνική συνάρτηση σφάλματος:

$$E_{lms} = \sum_{q,n,j} [nnp_j(n, q) - \delta(j, P(n, q))]^2 + \sum_{q,n,j} [nnd_j(n, q) - \delta(j, D(n, q))]^2 + \\ + \sum_{q,n,j} [nnc_j(n, q) - \delta(j, C(n, q))]^2$$

όπου q είναι ένας δείκτης από κομμάτια στο σύνολο εκπαίδευσης, n είναι ένας δείκτης από νότες μέσα σε ένα κομμάτι και j είναι ένας δείκτης από τονικότητες, διάρκειες και συνοδευτικές συγχορδίες στο στρώμα NNL. $P(n, q)$, $D(n, q)$, $C(n, q)$, είναι δείκτες για τις τονικότητες, διάρκειες και συνοδευτικές συγχορδίες στόχους αντίστοιχα, της νότας n , του κομματιού q . Τέλος, $\delta(a, b)$ είναι η συνάρτηση d του Kronecker, όπου

$$\delta(a, b) = \begin{cases} 1, & \text{αν } a = b \\ 0, & \text{αν } a \neq b \end{cases}$$

Αυτό το μέτρο ελαχιστοποιείται όταν οι έξοδοι από τις μονάδες που ανταποκρίνονται στις σωστές προβλέψεις είναι 1 και οι έξοδοι από τις άλλες μονάδες είναι 0.

Ένα επιπλέον μέτρο αποδοτικότητας στο περιεχόμενο των μονάδων εξόδου είναι το μέτρο της δεσμευμένης πιθανότητας όλες οι νότες ενός κομματιού από το σύνολο εκπαίδευσης να είναι σωστές. Αυτή η πιθανότητα εκφράζεται από το γινόμενο²

$$L = \prod_{q,n} nnlp_{P(n,q)}(n, q) \ nnld_{D(n,q)}(n, q) \ nnlc_{C(n,q)}(n, q)$$

Ένα άλλο κριτήριο είναι ο λογαριθμικός τύπος

$$E = -\log L = - \sum_{q,n} \log nnlp_{P(n,q)}(n, q) + \log nnld_{D(n,q)}(n, q) + \log nnlc_{C(n,q)}(n, q)$$

που είναι πιο εύκολο στη χρήση [10] και χρησιμοποιεί τα ίδια όρια με το παραπάνω.

Με τη μέθοδο της προς τα πίσω ολίσθησης του σφάλματος καθορίζονται τα βάρη στις συνδέσεις έτσι ώστε να ελαχιστοποιείται το E . Αυτό επιτυγχάνεται με τον υπολογισμό της κλίσης του E , δηλαδή με τον υπολογισμό των μερικών παραγώγων ως προς τα βάρη του στρώματος εξόδου, εισόδου και του στρώματος πλαισίου, δηλαδή $\frac{\partial E}{\partial \vec{W}}$, $\frac{\partial E}{\partial \vec{V}}$, $\frac{\partial E}{\partial \vec{U}}$ αντίστοιχα. Πρώτα θα πρέπει να υπολογίσουμε την κλίση συναρτήσει της δραστηριότητας στις μονάδες του στρώματος NND, και μετά οπισθοδρομούμε την κλίση για τον υπολογισμό των βαρών στα προηγούμενα στρώματα. Για τις μονάδες τονικότητας του στρώματος NND

$$\frac{\partial E}{\partial \overrightarrow{nndp}(n, q)} = \left[\frac{\overrightarrow{nndp}(n, q) - \vec{p}_P(n, q)}{d_P(n, q)} - \sum_i nnlp_i(n, q) \frac{\overrightarrow{nndp}(n, q) - \vec{p}_i}{d_i} \right]$$

Η μέθοδος της προς τα πίσω ολίσθησης, όμως, δεν μπορεί να εκπαιδεύσει το CONCERT απευθείας, μιας και περιέχει αναδραστικές συνδέσεις και η προαναφερθείσα μέθοδος εφαρμόζεται για την εκπαίδευση προωθητικών δικτύων. Πολλές διαφοροποιήσεις έχουν προταθεί για την χρήση της μεθόδου της προς τα πίσω ολίσθησης για αναδραστικά δίκτυα και για το CONCERT χρησιμοποιείται η προς τα πίσω ολίσθηση στο χρόνο (Back Propagation Through Time, BPTT) [57], η οποία μετατρέπει ένα αναδραστικό νευρωνικό σε ένα ισοδύναμο προωθητικό.

Χρησιμοποιήθηκε η μέθοδος του ξεδιπλώματος στον χρόνο με τόσες φορές όσα είναι και τα στοιχεία της χρονοσειράς. Για παράδειγμα, αν το κομμάτι στο οποίο εκπαιδεύεται το δίκτυο έχει 10 νότες τότε θα δημιουργηθεί ένα δίκτυο που να παίρνει τις εισόδους, $x(\vec{1}) \dots x(\vec{10})$, όπου το $x(\vec{t})$ παριστάνει το διάνυσμα που περιγράφει τη νότα της χρονικής στιγμής t . Εφαρμόζοντας την προς τα πίσω ολίσθηση του σφάλματος στο ισοδύναμο προωθητικό δίκτυο επιφέρει τις αλλαγές στα βάρη $\{\delta \vec{W}, \delta \vec{V}^{(i)}, \delta \vec{U}^{(i)}\}$, όπου το (i) δείχνει το 'αντίγραφο' του του αναδραστικού δικτύου στο οποίο γίνεται η αλλαγή, όπως φαίνεται στο σχήμα 2.2. Η τελική αλλαγή που επιφέρεται στα βάρη του δικτύου (υπενθυμίζουμε πως τα βάρη του προωθητικού δικτύου σε κάθε αντιγραφή (i) αντιστοιχούν σε ένα σύνολο βαρών) είναι

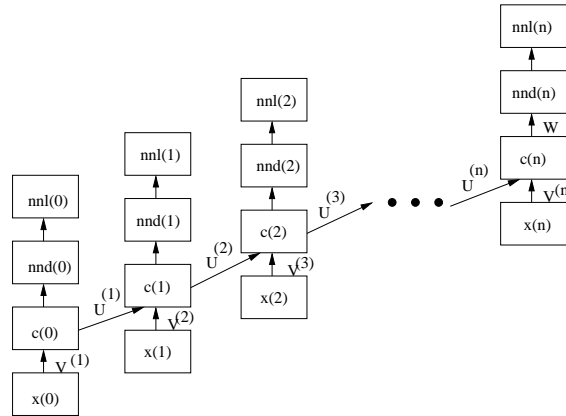
$$\delta \vec{W}$$

²Αυτή η πιθανότητα εκφράζει ανεξαρτησία μεταξύ των προβλέψεων των νωτών, κάτι το οποίο δεν ισχύει στο CONCERT. Παρόλα αυτά, ο παραπάνω τύπος αποτελεί μία λιγότερο διαισθητική προσέγγιση η οποία λειτουργεί ως μέτρο πληροφορίας.

$$\delta \vec{V} = \sum_{i=1}^n \delta \vec{V}^{(i)}$$

$$\delta \vec{U} = \sum_{i=1}^n \delta \vec{U}^{(i)}$$

και η τελική άθροιση συμβαίνει στο τέλος κάθε εποχής (βλ. Κεφάλαιο 1.6.1), δηλαδή αφού παρουσιαστεί όλη η χρονοσειρά.



Σχήμα 2.2: Η ξεδιπλωμένη στον χρόνο αρχιτεκτονική του CONCERT

Η παραπάνω μέθοδος για μεγάλες χρονοσειρές έχει μειονεκτήματα όχι μόνο από την άποψη του υπολογιστικού κόστους, αλλά και της ίδιας της διαδικασίας της εκπαίδευσης λόγω της εξαφανιζόμενης κλίσης. Οι χρονοσειρές όμως που χρησιμοποιούνται σε αυτά τα πειράματα είναι μέσα στα αριθμητικά επιτρεπτά πλαίσια.

Εξετάζοντας τη ξεδιπλωμένη αρχιτεκτονική γίνεται κατανοητό το πώς το CONCERT μπορεί να βρει συνοχή μεταξύ απομακρυσμένων στοιχείων της χρονοσειράς. Ας θεωρήσουμε για παράδειγμα την είσοδο στο πρώτο βήμα, δηλαδή για $t = 1$, $x(1)$, το οποίο είναι συνδεδεμένο με τη πρόβλεψη για το βήμα t , $nnl(t)$, μέσω των ενδιάμεσων στρωμάτων $c(1), c(2) \dots c(t)$ και $nnd(t)$. Τα βάρη κατά μήκος αυτής της διαδρομής προσαρμόζονται μέσω προς τα πίσω ολίσθησης έτσι ώστε αν το $x(1)$ έχει κάποιο ρόλο για τη πρόβλεψη του $nnl(t)$ αυτή να ληφθεί υπόψιν. Αυτή η ιδιότητα περιφρουρείται από το γεγονός ότι το πρώτο βήμα θα έχει λόγο στη πρόβλεψη καθόλη τη διάρκεια της χρονοσειράς μέσα από το δίκτυο κατά τη διαδικασία εκπαίδευσης. Αν δίναμε στο δίκτυο μικρότερο ορίζοντα, το ξεδιπλώναμε δηλαδή λιγότερες φορές από το μέγεθος της χρονοσειράς, τότε δε θα ήμασταν βέβαιοι πως το $x(1)$ θα είχε κάποιο ρόλο για τη πρόβλεψη κατά μήκος της συνολικής διάρκειας της χρονοσειράς.

2.5 Αναπαράσταση μουσικών αντικειμένων

Έχοντας περιγράψει την αρχιτεκτονική του CONCERT και τον τρόπο με τον οποίο εκπαιδεύεται, θα δούμε τον τρόπο με τον οποίον αναπαρίστανται τα μουσικά αντικείμενα. Τα τρία χαρακτηριστικά των αντικειμένων αυτών που πρέπει να αναπαρασταθούν είναι η τονικότητα, η διάρκεια και η αρμονική συνοδεία από συγχορδίες.

Η τονικότητα και η διάρκεια των νοτών είναι η *μελωδική γραμμή* και η αρμονική συνοδεία αποτελείται από συγχορδίες. Συνήθως οι συγχορδίες αλλάζουν σπανιότερα από τις νότες της μελωδικής γραμμής που συνοδεύουν και αυτό κωδικοποιείται με επανάληψη των συγχορδιών πάνω από νότες μέχρι να συμβεί η αλλαγή σε συγχορδιών. Παρακάτω αναλύουμε τα τρία αυτά χαρακτηριστικά.

2.5.1 Αναπαράσταση τονικότητας

Για να έχει την ικανότητα το CONCERT να αναπαράγει ένα εύρος ειδών μουσικής, χρειάζεται να έχει επαρκώς μεγάλο εύρος τονικότητας, περίπου τεσσάρων οκτάβων. Χρησιμοποιώντας τον συνήθη συμβολισμό των νοτών ακολουθούμενο από έναν αριθμό που υποδηλώνει το σχετικό ύψος της οκτάβας στην οποία βρίσκεται η νότα αυτή έχουμε τους εξής συμβολισμούς:

$$C1, D1, E1, \dots, C2, D2, \dots, C3, \dots, C4, \dots, A4, B4, C5$$

όπου $C1$ είναι η χαμηλότερη τονικότητα. Οι διέσεις και οι υφέσεις συμβολίζονται με $\#$ και b αντίστοιχα. Το συνολικό εύρος των νοτών λοιπόν, είναι 49 νότες, αν αναλογιστούμε πως κάθε οκτάβα περιέχει 12 χρωματικά βήματα.

Ο πιο απλός τρόπος για την αναπαράσταση της τονικότητας μάλλον είναι να υπάρχει μία μονάδα για κάθε τονικότητα. Για παράδειγμα, η νότα $C1$ να αναπαρίσταται με το διάνυσμα $[1\ 0\ 0\ \dots]^T$, η νότα $C2$ με το διάνυσμα $[0\ 1\ 0\ \dots]^T$ κ.ο.κ. Ένας άλλος τρόπος θα μπορούσε να είναι η αναπαράσταση όλων των νοτών από μία μονάδα, η οποία θα δίνει τιμές που ανάλογα με το εύρος τους θα αντιστοιχούν σε μία τονικότητα. Κάποιος θα μπορούσε να πει πως οποιοσδήποτε τρόπος είναι αποδεκτός, αφού τελικά η μόνη επίπτωση είναι στο τρόπο εκπαίδευσης από τη μέθοδο της προς τα πίσω ολίσθησης του σφάλματος [22]. Παρόλα αυτά, παραδείγματα από ερευνητές προτείνουν πως όσο πιο προσαρμοσμένη στο πρόβλημα είναι η αναπαράσταση των στοιχείων τόσο πιο εύκολη και αποτελεσματική είναι η δουλειά του αλγορίθμου εκπαίδευσης [13, 44].

Η προσέγγιση που κάνουμε για την αναπαράσταση νοτών, εκτός των προαναφερθέντων, έχει και βάσεις σε ψυχοακουστικές μελέτες [30]. Με αυτό εννοούμε πως το CONCERT κρίνει ως παρόμοιες, νότες που οι άνθρωποι κρίνουν παρόμοιες, έτσι ώστε οι αναπαράσταση των νοτών στο CONCERT να μοιάζει με την αναπαράσταση των νοτών στον ανθρώπινο εγκέφαλο. Για παράδειγμα, οι άνθρωποι αντιλαμβάνονται εύκολα ως όμοιες τις νότες $C1$ και $C\#1$ απότι τις $C1$ και $A4$.

Ο Shepard [60], με μία έρευνα που έκανε ρωτώντας ανθρώπους για το ποιες νότες αντιλαμβάνονται ως όμοιες, κατέληξε στο συμπέρασμα πως η αντίληψη δύο νοτών ως όμοιες μειώνεται εκθετικά όσο αυξάνεται η απόστασή τους. Για την αναπαράσταση της τονικότητας, ο Shepard προτείνει τη χρήση ενός πενταδιάστατου χώρου ο οποίος χωρίζεται σε τρεις υποχώρους: τον υπόχωρο του *ύψους τονικότητας* (Pitch Height, PH) μίας διάστασης, τον υπόχωρο του *χρωματικού κύκλου* (Chroma Cycle, CC) δύο διαστάσεων και τον υπόχωρο του *κύκλου των πέμπτων* (Circle of Fifths, CF) που έχει δύο διαστάσεις. Θα αναφερόμαστε σε αυτή την αναπαράσταση ως PHCCCF. Τα στοιχεία του ύψους τονικότητας καθορίζουν τον λογάριθμο της συχνότητας μιας τονικότητας. Στον κύκλο των πέμπτων, η καθαρή πέμπτη μιας τονικότητας θεωρείται η ακριβώς γειτονική νότα, αντίστροφα με τους δείκτες του ρολογιού³. Το πόσο κοντά

³Η φυσική πέμπτη είναι ένα διάστημα μεγάλη σημασίας. Η συχνотική αναλογία μιας νότας από

λοιπόν, είναι δύο νότες στον πενταδιάστατο χώρο PHCCCF, υπολογίζεται απλά με τον υπολογισμό της Ευκλείδειας απόστασής τους στον πενταδιάστατο αυτό χώρο.

τόνος	PH	CC						CF					
C1	-9.798	+1	+1	+1	-1	-1	-1	-1	-1	-1	+1	+1	+1
F#1	-7.349	-1	-1	-1	+1	+1	+1	+1	+1	+1	-1	-1	-1
G2	-2.041	-1	-1	-1	-1	+1	+1	-1	-1	-1	-1	+1	+1
C3	0	+1	+1	+1	-1	-1	-1	-1	+1	+1	+1	+1	+1
D#3	1.225	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1
E3	1.633	-1	+1	+1	+1	+1	+1	+1	-1	-1	-1	-1	-1
A4	8.573	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
C5	9.798	+1	+1	+1	-1	-1	-1	-1	-1	-1	+1	+1	+1
παύση	0	+1	-1	+1	-1	+1	-1	+1	-1	+1	-1	+1	-1

Πίνακας 2.2: Αναπαράσταση PHCCCF για επιλεγμένες τονικότητες.

τόνος						
C	-1	-1	-1	-1	-1	-1
C#	-1	-1	-1	-1	-1	+1
D	-1	-1	-1	-1	+1	+1
D#	-1	-1	-1	+1	+1	+1
E	-1	-1	+1	+1	+1	+1
F	-1	+1	+1	+1	+1	+1
F#	+1	+1	+1	+1	+1	+1
G	+1	+1	+1	+1	+1	-1
G#	+1	+1	+1	+1	-1	-1
A	+1	+1	+1	-1	-1	-1
A#	+1	+1	-1	-1	-1	-1
B	+1	-1	-1	-1	-1	-1

Πίνακας 2.3: Αναπαράσταση της τονικότητας στον χρωματικό κύκλο CC.

Θα αναφερθούμε στα θετικά της αναπαράστασης PHCCCF σε σχέση με μια σειριακή, συχνοτική αναπαράσταση. Ας θεωρήσουμε τον τρισδιάστατο υπόχωρο που ορίζεται από τους υποχώρους PH και CC. Μπορούμε να σκεφτούμε αυτόν τον υπόχωρο ως μία έλικα που άνεβοκατεβαίνει μέσα στον χώρο. Το ύψος που βρίσκεται αυτή η έλικα καθορίζει την τιμή στον άξονα PH, ενώ η γωνία η οποία σχηματίζει η έλικα καθορίζει τις τιμές που θα έχει ο χώρος CC. Εξετάζοντας λοιπόν, το παρακάτω απλό παράδειγμα, βλέπουμε πως οι νότες $C1$ και $C2$ βρίσκονται πιο κοντά απότι οι νότες $C1$ και $B1$, αν και το δεύτερο ζευγάρι βρίσκεται πιο κοντά συχνοτικά απότι το δεύτερο.

Η προσθήκη του CF στον παραπάνω τρισδιάστατο χώρο, προικίζει την αναπαράσταση με επιπλέον επιθυμητές ιδιότητες. Πρώτον, ο κύκλος των πέμπτων εντοπίζει τις νότες σε ένα μουσικό κλειδί. Για παράδειγμα, αν πάρουμε το κλειδί της C μείζονος, ή ισοδύναμα της A ελάσσονος, βλέπουμε πως όλες οι διατονικές νότες, $C D E F G A B$, είναι τοπικά ομαδοποιημένες στον κύκλο των πέμπτων. Ομοίως

την φυσική πέμπτη της είναι 2:3, όπως της οκτάβας είναι 1:2.

βαθμός	τόνος	απόσταση	βαθμός	τόνος	απόσταση
1	D2	2.528	6	C#2	4.422
2	E2	2.779	7	A2	4.422
3	A1	3.399	8	E1	4.411
4	B1	3.859	9	G1	4.497
5	C2	4.130	10	G2	4.497

Πίνακας 2.4: Οι δέκα μικρότερες αποστάσεις της αναπαράστασης $A1 D2 E2$.

βαθμός	τόνος	απόσταση	βαθμός	τόνος	απόσταση
1	D2	2.373	6	D#2	3.774
2	C2	3.277	7	A1	3.946
3	E2	3.538	8	F2	4.057
4	C#2	3.654	9	A#1	4.146
5	B1	3.714	10	G1	4.323

Πίνακας 2.5: Οι δέκα μικρότερες αποστάσεις της αναπαράστασης $A1 D2 D\#2$.

ομαδοποιούνται και οι νότες σε πεντατονικές κλίμακες. Δεύτερον, και μάλλον κρισιμότερο, ο κύκλος των πέμπτων εξηγεί την υποκειμενική ισοτιμία στα διαστήματα σε μία διατονική κλίμακα. Ο Shepard διεπίστωσε πως ο άνθρωπος έχει την τάση να αντιλαμβάνεται τα διαδοχικά διαστήματα μιας διατονικής κλίμακας ως ίσα παρόλο που μερικά διαστήματα είναι τα μισά από άλλα. Για παράδειγμα, στην κλίμακα C ματζόρε τα διαστήματα $E - F$ και $B - C$ είναι μισού τόνου, ενώ όλα τα άλλα είναι ολόκληρου. Ένας συνδυασμός του ύψους τονικότητας και του κύκλου των πέμπτων μας επιτρέπει μία αναπαράσταση όπου οι μικρές και οι μεγάλες δεύτερες απέχουν το ίδιο. Αυτό επιτυγχάνεται με τη χρήση μιας αναλογίας 3:1 του χρωματικού κύκλου προς τον κύκλο των πέμπτων.

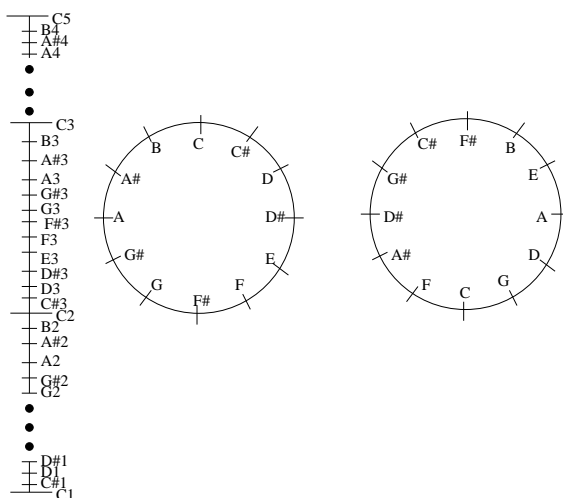
Άλλη μία επιθυμητή ιδιότητα της αναπαράστασης $PHCCCF$ είναι πως οι αποστάσεις των τονικοτήτων παραμένουν αναλλοίωτες από μετασχηματισμούς. Για παράδειγμα η απόσταση των $D2$ και $G\#4$ είναι η ίδια με την απόσταση των $E2$ και $A\#4$.

Η σχετική σημαντικότητα στο ρόλο των στοιχείων PH , CC και CF στην επιλογή της νότας, μπορεί να καθοριστεί από την διάμετρο των κύκλων των CC και CF . Για παράδειγμα, αν οι διάμετροι των κύκλων είναι ίδιες, τότε η απόσταση των νοτών C και G είναι ίδια με την απόσταση των C και B . Αυτό συμβαίνει γιατί η B απέχει ένα 'δόντι' από την C στο γρανάζι του χρωματικού κύκλου και πέντε 'δόντια' στο γρανάζι του κύκλου των πέμπτων, ενώ η C απέχει από την G τις αντίστροφες αποστάσεις στα αντίστοιχα γρανάζια. Αν αυξηθεί η διάμετρος του CC , τότε η C έρχεται πιο κοντά στην B απότι στη G ενώ αν μειωθεί συμβαίνει το αντίθετο. Αν, από την άλλη, μειωθεί η διάμετρος και των δύο κύκλων κρατώντας σε σταθερό μήκος την ευθεία του ύψους τονικότητας, τότε οι κύκλοι των CC και CF παίζουν μικρότερο ρόλο στον καθορισμό των αποστάσεων μεταξύ νοτών. Ο Shepard υποστηρίζει πως η κλίμακα μεγεθών που έχει ο κάθε άνθρωπος στο μυαλό του εξαρτάται από την μουσική εμπειρία που έχει. Με αυτό σαν δεδομένο υποστηρίζει πως μια λογική αναπαράσταση για κάποιον έμπειρο μουσικό είναι αναπαράσταση σύμφωνα με την οποία το μήκος της διαμέτρου των CC και CF είναι το ίδιο, και μάλιστα είναι ίσο με το μήκος μίας οκτάβας στο ευθύγραμμο τμήμα του PH . Στο σχήμα 2.3 έχει τηρηθεί αυτή η κλίμακα αναπαράστασης.

Το τελευταίο ζήτημα που θα μας απασχολήσει είναι η μορφοποίηση της αναπαράστασης PHCCCF σε ένα διάγραμμα που θα επιδρά στις συνάψεις του νευρωνικού μας δικτύου. Η ευθεία προσέγγιση είναι η χρήση πέντε μονάδων, μία για το ύψος τονικότητας και από δύο για τον χρωματικό κύκλο και τον κύκλο των πέμπτων⁴.

Ένα πρόβλημα που προκύπτει με αυτή τη προσέγγιση είναι το ότι αν οι μονάδες έχουν τη συνήθη σιγμοειδή συνάρτηση ενεργοποίησης, η ισότητα στις αποστάσεις νοτών στο ύψος τονικότητας ή και στους κύκλους στο χώρο δραστηριότητας της κάθε μονάδας δεν διατηρείται στο χώρο των μονάδων εισόδου. Αυτό σημαίνει πως οι μονάδες πλαισίου που επιχειρούν να ενεργοποιήσουν τις μονάδες του στρώματος NND δεν αποδίδουν όλα τα θετικά της αναπαράστασης. Ένα δεύτερο πρόβλημα με το απλό πενταδιάστατο πλάνο είναι πως κάθε μονάδα κωδικοποιεί κάθε συντεταγμένη ευθέως, πράγμα που θα οδηγούσε στο ότι μικρές διαταραχές στο διάγραμμα δραστηριότητας θα οδηγήσουν σε μεγάλες παρερμηνείες.

Εξαιτίας αυτών των προβλημάτων, υιοθετείται μία εναλλακτική αναπαράσταση των στοιχείων των δύο κύκλων. Αυτή η αναπαράσταση εκφράζεται με έξι μονάδες που δέχονται δυαδικές τιμές για να αναπαραστήσουν τον κάθε τόνο στον κάθε κύκλο. Στον πίνακα 2.3 παρουσιάζουμε την αναπαράσταση του χρωματικού κύκλου. Με αυτόν τον τρόπο διατηρούνται οι στοιχειώδεις αποστάσεις στις νότες του χρωματικού κύκλου: η απόσταση μεταξύ δύο νοτών είναι μονοτονικά εξαρτημένη από την γωνία που σχηματίζουν οι νότες αυτές. Επίσης, μιας και κάθε μονάδα έχει κωδικοποιηθεί με δύο διακριτές αξίες, είναι λιγότερο ευαίσθητες σε θόρυβο από μία αναπαράσταση που θα μετρούσε δύο μη διακριτές συνιστώσες.



Σχήμα 2.3: Η αναπαράσταση PHCCCF

Δυστυχώς, δεν υπάρχει παρόμοιος τρόπος για να κωδικοποιήσουμε το ύψος τονικότητας, μία προσπάθεια να γίνει αυτό σε έναν χώρο Boole θα οδηγούσε σε πίνακες

⁴Ο αναγνώστης μπορεί να απορήσει για ποιο λόγο χρειαζόμαστε δύο διαστάσεις για τον καθορισμό σημείων πάνω σε κύκλο σταθερής διαμέτρου, από τη στιγμή που μόνο το όρισμα της γωνίας θα μπορούσε να μας καθορίσει πλήρως τη θέση του. Το ζήτημα όμως είναι ότι εμείς αναζητούμε την απόσταση των στοιχείων αυτών στον ευκλείδειο χώρο που είναι εμβαπτισμένα. Αν δηλαδή, κόβαμε σε ένα σημείο έναν από τους κύκλους, τότε οι γειτονικές νότες εκατέρωθεν του σημείου κοπής θα ήταν πλέον οι πιο απομακρυσμένες.

υπερβολικά μεγάλων διαστάσεων. Γιαυτό το λόγο χρησιμοποιούμε μία γραμμική μονάδα για την αναπαράσταση του ύψους τονικότητας. Οι διαφορετικές τιμές που πρέπει να λαμβάνονται είναι 49, μπορούμε όμως να μην απαιτούμε μεγάλη ακρίβεια για την εμφάνιση του αποτελέσματος από αυτή τη μονάδα. Αυτό γιατί η μονάδα του ύψους τονικότητας μας δίνει κυρίως πληροφορίες για το ύψος της οκτάβας της νότας που είναι υποψήφια να ακολουθήσει και για την ακριβή της τονικότητα κυριότερο ρόλο παίζουν οι υπόλοιπες διαστάσεις. Για την δραστηριότητα αυτής της μονάδας επιλέχθηκε μία κλίμακα με 49 ισαπέχουσες νότες εύρους από το -9.798 για την $C1$ ως το 9.798 για την $C5$. Αυτή η κλίμακα επιτυγχάνει τις επιθυμητές ιδιότητες που αναφέραμε προηγουμένως που αφορούν τις αποστάσεις στο ύψος τονικότητας και τις διαμέτρους των κύκλων.

Η αναπαράσταση PHCCCF αποτελείται από 13 μονάδες συνολικά. Στον πίνακα 2.2 βλέπουμε παραδείγματα των μορφών δραστηριότητας για κάποιες νότες. Οι παύσεις ακολουθούν μια μοναδική κωδικοποίηση εντελώς διαφορετική από τις άλλες νότες. Το τέλος του κάθε κομματιού κωδικοποιείται με την παράθεση συνεχόμενων παύσεων.

Μη ξεχνάμε όμως, πως το στρώμα NND πρέπει να δείχνει διάφορες υποψήφιες τονικότητες, άρα θα πρέπει να κωδικοποιεί επιπλέον εναλλακτικές. Αν, για παράδειγμα οι νότες $A1$, $D2$ και $E3$ είχαν την ίδια πιθανότητα να είναι οι επόμενες νότες, το στρώμα NND θα έπρεπε να υποδεικνύει και τις τρεις εναλλακτικές προτάσεις. Για να το πετύχει αυτό πρέπει να παράγει ένα διάνυσμα δραστηριότητας που να είναι πιο κοντά μόνο στις τρεις πιθανότητες \vec{p}_{A1} , \vec{p}_{D2} και \vec{p}_{E2} . Το σημείο στον πενταδιάστατο χώρο του PHCCCF που είναι πιο κοντά και στα τρία αυτά σημεία είναι το διάνυσμα μέσος όρος αυτών των διανυσμάτων, δηλαδή το διάνυσμα $\frac{\vec{p}_{A1} + \vec{p}_{D2} + \vec{p}_{E2}}{3}$. Στον πίνακα 2.4 φαίνεται πως κάτι τέτοιο ισχύει, αν θεωρήσουμε όμως το παράδειγμα των νοτών $A1$, $D2$ και $D\#2$, τότε, απ' ότι φανερώνει ο πίνακας 2.5, δεν μπορούμε να εμπιστευτούμε το διάνυσμα του μέσου όρου. Φαίνεται δηλαδή, πως ορισμένες ομάδες από νότες είναι πιο σφιχτές από άλλες στον χώρο της αναπαράστασης PHCCCF. Απ' αυτό συμπεραίνουμε πως η αναπαράσταση PHCCCF όχι μόνο μας συνιστά μία ομαδοποίηση ομοιότητας για τις νότες, αλλά παράγει και έναν περιορισμό για το ποιος συνδυασμός νοτών μπορεί να προταθεί σε ένα δεδομένο βήμα. Σαν γενική έννοια οι αυθαίρετοι περιορισμοί δεν είναι θεμιτό να υπάρχουν, εδώ όμως είναι θεωρητικά υποκινούμενοι.

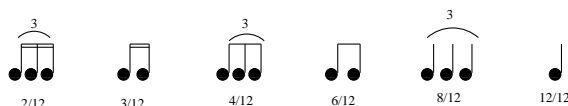
Ένα σοβαρό μειονέκτημα της αναπαράστασης PHCCCF είναι ότι βασίζεται στην ομοιότητα ομάδων νοτών που βρίσκονται σε απομόνωση. Οι άνθρωποι δεν αντιλαμβάνονται απομονωμένες νότες αλλά νότες που ανήκουν σε ένα μουσικό περιεχόμενο που καθορίζει μία κλίμακα [29, 32, 33]. Το CONCERT θα μπορούσε να βελτιωθεί με μία προσθήκη τέτοια ώστε να λαμβάνει υπόψιν και τη μουσική κλίμακα στην οποία κινείται μία μελωδική γραμμή.

2.5.2 Αναπαράσταση διάρκειας

Παρόλο που υπάρχουν μελέτες στο πώς οι άνθρωποι αντιλαμβάνονται την διάρκεια των νοτών [18, 25], δεν έχει υπάρξει μία αναπαράσταση των νοτών ανάλογη με αυτή του Shepard για την τονικότητα, αν και ο ίδιος ο Shepard προτείνει πως κάτι τέτοιο θα έπρεπε να συμβαίνει. Η αναπαράσταση της τονικότητας με την οποία λειτουργεί το CONCERT είναι προσπάθεια υιοθέτησης της παραπάνω πρότασης.

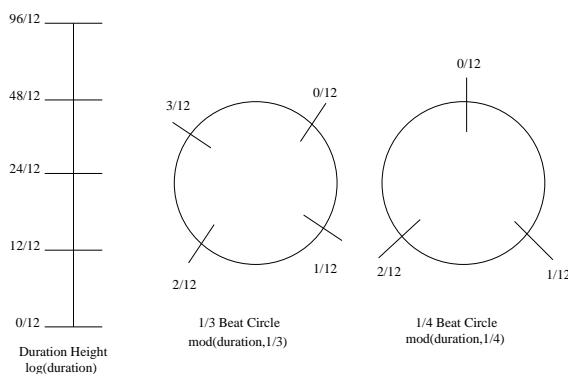
Η αναπαράσταση της διάρκειας βασίζεται στη διαίρεση του κάθε τετάρτου σε 12

μέρη. Πράγμα που σημαίνει πως το τέταρτο έχει αξία $\frac{12}{12}$, το ογδόο έχει αξία $\frac{6}{12}$, το τρίηχο ογδόων έχει αξία $\frac{4}{12}$ κ.ο.κ. Χρησιμοποιώντας αυτόν τον τρόπο υποδιαίρεσης μπορούμε να κατασκευάσουμε έναν πενταδιάστατο χώρο που αποτελείται από τρία στοιχεία, όπως φαίνεται στο σχήμα 2.5. Σε αυτήν την αναπαράσταση, κάθε διάρκεια καθορίζει ένα σημείο στο μονοδιάστατο υπόχωρο του ύψους διάρκειας, δύο συντεταγμένες στο δισδιάστατο χώρο του $1/3$ κύκλου και άλλες δύο συντεταγμένες στο δισδιάστατο χώρο του $1/4$ κύκλου. Το ύψος διάρκειας είναι ανάλογο του λογαρίθμου της διάρκειας, με τη λογαριθμική αυτή μετατροπή να ακολουθεί τον ψυχοφυσικό νόμο του Fechner που σχετίζει την διεγερτική ένταση με το λαμβανόμενο συναίσθημα. Το νόημα με τον κύκλο $1/n$ είναι η διάρκεια, αφού αφαιρέσουμε το μεγαλύτερο ακέραιο πολλαπλάσιο από το $1/n$. Για παράδειγμα, η αξία $18/12$ παριστάνεται με το σημείο $2/12$ στον κύκλο του $1/3$ και με το σημείο $0/12$ στον κύκλο του $1/4$. Έτσι, οι αξίες των τετάρτων και των ογδόων έχουν την ίδια αξία στον κύκλο του $1/4$, ενώ τα τρίηχα τετάρτων και τα τρίηχα ογδόων έχουν την ίδια αξία στον κύκλο του $1/3$ και τα τέταρτα και τα μισά έχουν ίδιες τιμές και στους δύο κύκλους.



Σχήμα 2.4: Κλασματική αναπαράσταση νοτών βάσει της μοντελοποίησής μας.

Ο πενταδιάστατος χώρος που αναλύσαμε παραπάνω κωδικοποιείται απευθείας με πέντε μονάδες. Οι υπόχωροι των κύκλων του $1/3$ και του $1/4$ δεν καθίσταται ανάγκη να κωδικοποιηθούν σε δυαδικούς χώρους, όπως ο χρωματικός κύκλος και ο κύκλος των πέμπτων της τονικότητας, μιας και τα στοιχεία τους είναι αρκετά απομακρυσμένα ώστε να δημιουργηθούν παρερμηνείες από λάθη.



Σχήμα 2.5: Μοντελοποίηση για την αναπαράσταση της διάρκειας.

Αξίζει να αναφέρουμε και άλλες προσεγγίσεις για την αναπαράσταση της διάρκειας. Μία ευθεία προσέγγιση αυτού του ζητήματος είναι η αναπαράσταση του χρόνου ευθέως, με την παρουσίαση της νότας προς αναπαράσταση τόσες φορές στο στρώμα εξόδου, όσες φορές επιτάσσει η αξία της. Για παράδειγμα, αν κρατήσουμε τον παραπάνω κατακερματισμού του τετάρτου σε 12 μέρη, για να εμφανιστεί μία

νότα αξίας μισής νότας, τότε θα έπρεπε να εμφανιστεί 24 φορές. Αυτή η προσέγγιση, η οποία ακολούθησε και ο Todd στην προσέγγισή του [63], έχει το σημαντικό μειονέκτημα του μεγάλου χώρου αποθήκευσης που χρειάζεται για μικρές πληροφορίες, πράγμα που κάνει ακόμα πιο δύσκολο το έργο του νευρωνικού δικτύου για κατανόηση κανονικοτήτων απομακρυσμένων νοτών.

2.5.3 Αναπαράσταση συγχορδιών

Η αναπαράσταση συγχορδιών βασίζεται στην ψυχοακουστική προσέγγιση των Laden και Keefe [30]. Παρακάτω αναλύουμε πρώτα την προσέγγιση αυτή και μετά εξηγούμε τις τροποποιήσεις που χρειάζεται για να γίνει επεξεργάσιμη από το CONCERT.

Οι συγχορδίες που χρησιμοποιούνται εδώ είναι στην θέση τη τονικής τους (δεν χρησιμοποιούνται δηλαδή αναστροφές τους) και κάθε συγχορδία αποτελείται από τρεις ή τέσσερις νότες. Μερικά παραδείγματα βλέπουμε στον πίνακα 2.6. Στα πνευστά όργανα και στην ανθρώπινη φωνή, η παραγωγή μίας νότας θα έχει σαν αποτέλεσμα την παραγωγή της θεμελιώδους συχνότητας της τονικότητας (π.χ. A3 με συχνότητα 440 Hz) αλλά και ενός αρμονικού φάσματος από συχνότητες που είναι ακέραια πολλαπλάσια της θεμελιώδους αυτής συχνότητας (880 Hz, 1320 Hz, 1760 Hz και 2200 Hz). Αν αντιστοιχίσουμε τις συχνότητες αυτές σε νότες, τότε θα έχουμε τις ακόλουθες αντιστοιχίες: 440 στην A3, 880 στην A4, 1320 στην E3, 1760 στην A5 και 2200 στην C#5. Έτσι, η αναπαράσταση της συγχορδίας εναποτίθεται στην αναπαράσταση των νοτών που την αποτελούν. Το επιθυμητό εύρος είναι από την C3 ως και την C7, χρειαζόμαστε δηλαδή 49 νότες. Τα είδη συγχορδιών που θα χρειαζόμαστε είναι οι ματζόρε, οι μινόρε και οι ελαττωμένα.

συγχορδίες	νότες			
C μείζονα	C3	E3	G3	
C ελάσσονα	C3	Eb3	G3	
C αυξημένη	C3	E3	G#3	
C ελαττωμένη	C3	Eb3	Gb3	
C7	C3	Eb3	G3	Bb3

Πίνακας 2.6: Παράδειγμα συγχορδιών και οι νότες από τις οποίες αποτελούνται.

Για την είσοδο και επεξεργασία των συγχορδιών στο CONCERT έγιναν διάφορες τροποποιήσεις. Η πληροφορία της οκτάβας δεν χρησιμοποιείται, μειώνοντας την διάσταση της αναπαράστασης από 49 σε 12. Η βαρύτητα της κάθε αρμονικής μειώνεται εκθετικά με την αρμονική σειρά της. Η πρώτη αρμονική κωδικοποιείται με βαρύτητα 1.0, η δεύτερη με 0.5, η τρίτη με 0.25 κ.ο.κ. Οι κλίμακα των δραστηριοτήτων άλλαξε εύρος, από 0 ως 1 έγινε από -1 ως 1. Τέλος, χρησιμοποιήθηκε μία επιπλέον παράμετρος που βασίζεται σε ψυχοακουστικές μελέτες και έχει να κάνει με την αντίληψη των συγχορδιών της τονικής, της υποδεσπόζουσας και της δεσπόζουσας ως ίδιες [28]. Προστέθηκε, λοιπόν, ένα παραπάνω στοιχείο για να φέρει αυτές τις συγχορδίες πιο κοντά. Το στοιχείο αυτό είχε τιμή +1.5 για τις τρεις αυτές συγχορδίες και -1.5 για όλες τις υπόλοιπες.

2.6 Βασικά αποτελέσματα εξομοιώσεων

Στην σύνθεση μουσικής με νευρωνικά δίκτυα υπάρχει κίνδυνος για αποδοχή του αποτελέσματος ως ικανοποιητικού, χωρίς το ενδιαφέρον για την απόδοση του νευρωνικού δικτύου. Αυτό συμβαίνει γιατί το νέο μουσικό περιεχόμενο που παράγει το νευρωνικό δίκτυο, αποτελεί κάτι το ενδιαφέρον μερικές φορές, χωρίς να είναι απαραίτητα σωστό, από την άποψη ότι μπορεί να είναι πιο τυχαίο και να μην στηρίζεται και τόσο στο περιεχόμενο στο οποίο εκπαιδεύτηκε. Αυτό φαίνεται από το γεγονός ότι καμιά φορά, τυχαία σειρά νοτών παιγμένες σε ένα πιάνο, για παράδειγμα, μπορεί να παράγει τυχαίες μελωδίες που σε μερικούς ακούγονται ενδιαφέρουσες. Έτσι λοιπόν ξεκινάμε την εξέταση των αποτελεσμάτων του CONCERT με βάση κάποιες απλές τεχνητές μελωδίες που περιέχουν μόνο εναλλαγές στις νότες, άρα οι διάρκειες και οι συγχορδίες αγνοούνται.

2.6.1 Επέκταση της κλίμακας C μείζονα

Για να αρχίσουμε με ένα απλό πείραμα, το CONCERT εκπαιδεύτηκε σε μία απλή σειρά νοτών που αποτελείται από τρεις οκτάβες της κλίμακας C μείζονος $C1 D1 \dots B3$. Ο στόχος σε κάθε βήμα ήταν η επόμενη νότα της κλίμακας. Το CONCERT θεωρούμε πως εκπαιδεύτηκε αν σε κάθε βήμα η μονάδα που παριστάνει την επόμενη νότα στο στρώμα NNL είναι πιο δραστήρια από τις άλλες. Μετά την εκπαίδευση, το CONCERT δοκιμάστηκε στο να παράγει την τέταρτη οκτάβα της παραπάνω κλίμακας. Το παραπάνω πείραμα έγινε 10 φορές με αρχικά τυχαία βάρη στις συνάψεις, 15 εσωτερικούς νευρώνες, βήμα εκπαίδευσης 0.005 και το αποτέλεσμα ήταν ότι και τις 10 φορές το CONCERT αναπαρήγαγε σωστά τις νότες της κλίμακας στην τέταρτη οκτάβα, μόνο που στις 4 από τις 10 η τελευταία νότα ήταν μία οκτάβα χαμηλότερα απότι έπρεπε, δηλαδή ήταν η $C4$ αντί για την $C5$. Στον πίνακα 2.7 φαίνονται οι τρεις πιο πιθανές επόμενες νότες στο στρώμα NNL, μαζί με την πιθανότητα που τις χαρακτηρίζει, για τις οκτάβες 3 και 4. Οι οκτάβα 3 ανήκει στο σύνολο εκπαίδευσης, ενώ η 4 αποτελεί νέο περιεχόμενο. Παρατηρούμε ότι οι πιθανότητες που χαρακτηρίζουν τις επόμενες νότες είναι με μεγάλη βεβαιότητα οι σωστές.

Η εκπαίδευση του επιτεύχθηκε με 2 μονάδες πλαισίου, ενώ αν βάζαμε περισσότερες μονάδες πλαισίου μεγάλωνε η ικανότητα γενίκευσης. Δοκιμάστηκε και εκπαίδευση σε σύνολο δύο οκτάβων, όμως το αποτέλεσμα ήταν μερικές φορές να προβλέπει νότα χαμηλότερης οκτάβας.

2.6.2 Εκπαίδευση στη δομή των διατονικών κλιμάκων

Το CONCERT δοκιμάστηκε με εκπαίδευση σε διατονικές κλίμακες σε διάφορα κλειδιά, σε σύνολο εκπαίδευσης εύρους μίας οκτάβας, για παράδειγμα η D μείζονα $D1 E1 F\#1 G1 A1 B1 C\#2 D2$. Από τις 37 διατονικές κλίμακες που μπορούν να κατασκευαστούν από το δεδομένο εύρος, $C1 - C5$, χρησιμοποιήθηκαν οι 28 σαν σύνολο εκπαίδευσης και οι 9 σαν σύνολο δοκιμής. Από τις 10 διαφορετικές δοκιμές με τυχαία αρχικοποίηση των βαρών και 20 κρυφούς νευρώνες, το CONCERT μπορούσε να αναπαράγει το σύνολο εκπαίδευσης σε περίπου 55 περάσματα. Όσον αφορά τη γενίκευση, το CONCERT δοκιμάστηκε για κάθε δεδομένη νότα σε καινούρια κλί-

μακα⁵ και είχε ποσοστό επιτυχίας 98.4%. Τα λάθη που έγιναν ήταν μετατροπή της νότας μία οκτάβα ή μισό τόνο μακριά από το αναμενόμενο.

είσοδος	έξοδος		
C3	D3 0.961	C3 0.017	E3 0.014
D3	E3 0.972	D3 0.012	F3 0.007
E3	F3 0.982	D#3 0.008	G3 0.006
F3	G3 0.963	F3 0.015	A3 0.010
G3	A3 0.961	G3 0.024	B3 0.012
A3	B3 0.972	A3 0.025	C4 0.002
B3	C4 0.979	A#3 0.010	C#4 0.005
C4	D4 0.939	C4 0.040	E4 0.009
D4	E4 0.968	D4 0.018	F4 0.006
E4	F4 0.971	D#4 0.016	E3 0.005
F4	G4 0.931	F4 0.037	F#4 0.015
G4	A4 0.938	G4 0.044	B4 0.007
A4	B4 0.915	A4 0.080	A#4 0.003
B4	C5 0.946	A#4 0.040	B#4 0.011

Πίνακας 2.7: Επιδόσεις στις οκτάβες 3 και 4 της C μείζονος διατονικής κλίμακας.

Δοκιμάστηκε και μία έκδοση του CONCERT που δεν χρησιμοποιεί την αναπαράσταση RHCCCF, αλλά την αναπαράσταση μόνο του ύψους τονικότητας στο στρώμα NND, όμως το καλύτερο ποσοστό ευστοχίας που κατάφερε ήταν 54.4%.

2.6.3 Εκπαίδευση σε τυχαίο περίπατο

Σαν τυχαίο περίπατο θεωρούμε την ακολουθία από νότες εντός μίας κλίμακας που δύο διαδοχικές νότες απέχουν μία διατονική νότα (το αν η επόμενη νότα θα είναι χαμηλότερη ή ψηλότερη καθορίζεται τυχαία). Στο πείραμα αυτό έχουμε 100 τυχαίους περιπάτους αποτελούμενους από 10 νότες και το CONCERT εκπαιδεύτηκε σε αυτούς για 50 περάσματα με 15 εσωτερικούς νευρώνες. Η δοκιμή έγινε με 100 νέους τυχαίους περιπάτους. Σε κάθε τυχαίο περίπατο η επαλήθευση έγινε με τη σύγκριση των δύο νοτών με τις μεγαλύτερες πιθανότητες που έδινε το CONCERT ,με την επόμενη νότα του περιπάτου. Τα ποσοστά επιτυχίας ήταν 99.95%, ενώ αν εφαρμόσουμε το παραπάνω πείραμα σε τυχαίο περίπατο στη χρωματική κλίμακα το ποσοστό επιτυχίας είναι 100%.

2.6.4 Εκπαίδευση σε διεσπαρμένο τυχαίο περίπατο

Διεσπαρμένο τυχαίο περίπατο εννοούμε μία σειρά 10 νοτών της μορφής

$$a_1, b_1, a_2, b_2, \dots, a_5, b_5$$

, όπου a_i και b_i είναι τυχαία επιλεγμένες νότες, και η νότα a_{i+1} είναι μία διατονική νότα πάνω ή κάτω από τη νότα a_i στην κλίμακα μείζονος, ομοίως για τις b_{i+1} και την b_i . Το σύνολο εκπαίδευσης ήταν 200 τέτοιοι περιπάτοι και το σύνολο δοκιμής ήταν

⁵Αυτός ο τρόπος δεν είναι ισοδύναμος με την πλήρη συνθετική λειτουργία του CONCERT, μιας και η είσοδος της επόμενης νότας δεν είναι η έξοδος του, αλλά η δεδομένη επόμενη νότα της κλίμακας.

άλλοι 100, ενώ το CONCERT είχε 25 εσωτερικούς νευρώνες. Είναι όμως απίθανο να γίνει πρόβλεψη της νότας b_i χρησιμοποιώντας μόνο σαν κριτήριο την νότα a_i αφού η συγκεκριμένη διαδοχή γίνεται τυχαία. Γιαυτό το λόγο, το CONCERT πρέπει να εντοπίσει ομοιότητες δομής χρησιμοποιώντας δύο προηγούμενες νότες. Έτσι, φαίνεται πως πρέπει να χρησιμοποιήσουμε πίνακες μετάβασης δεύτερης τάξης, με το κόστος όμως να μην έχουμε την επιθυμητή γενίκευση. Με την κατάλληλη τροποποίηση, ώστε το CONCERT να υπολογίζει πίνακες μετάβασης δεύτερης τάξης, είχαμε επιτυχία σε ποσοστό 67.1%.

2.6.5 Εκπαίδευση στις φράσεις με δομή AABA

Η φράση με δομή AABA αποτελείται από δύο ξεχωριστές φράσεις, την A και την B, παιγμένες με τη σειρά AABA. Για την εκπαίδευση του CONCERT σε αυτή τη δομή χρησιμοποιήθηκαν δύο χρωματικές φράσεις 5 νοτών με τυχαίο σημείο εκκίνησης, οι οποίες αντιπροσωπεύουν τις επί μέρους φράσεις A και B, και το τέλος του κύκλου AABA υποδεικνυόταν με μία παύση. Η κάθε μελωδία του συνόλου εκπαίδευσης, δηλαδή, αποτελείται από 21 νότες (λαμβάνοντας υπόψιν και την παύση). Παρακάτω βλέπουμε ένα παράδειγμα μιας τέτοιας φράσης.

F#2 G2 G#2 A2 A#2 F#2 G2 G#2 A2 A#2

C4 C#4 D4 D#4 E4

F#2 G2 G#2 A2 A#2 REST

Αυτές οι μελωδίες είναι απλά παραδείγματα μελωδιών που έχουν ταυτόχρονα 'λεπτή' αλλά και 'τραχεία' δομή. Η πρώτη φαίνεται στη δομή των επί μέρους φράσεων και η δεύτερη στη σχέση μεταξύ των φράσεων αυτών. Αυτά τα παραδείγματα κατασκευάστηκαν για να δούμε κατά πόσο το CONCERT μπορεί να εντοπίσει δομές πολλαπλών στρωμάτων και κατά πόσο μπορεί να βρει μακροπρόθεσμες συσχετίσεις, γεγονός τα οποία συναντάμε στην πραγματική μουσική.

Με εκπαίδευση του CONCERT σε 200 παραδείγματα και δοκιμή σε 100 και με αρχιτεκτονική που να περιλαμβάνει 35 κρυφούς νευρώνες, περνώντας το κάθε παράδειγμα 300 φορές για την προσαρμογή των βαρών, η απόδοση ήταν 97.3% για την αναγνώριση της τοπικής δομής, δηλαδή για την εύρεση της επόμενης νότας σε μία φράση, ενώ για την ολική δομή 58.4%.

2.6.6 Συζήτηση των μέχρι τώρα αποτελεσμάτων

Το CONCERT έδειξε ότι δίνει καλύτερα αποτελέσματα, όσον αφορά την τοπική δομή, από ότι ένας πίνακας μετάβασης πρώτης τάξης και απότι η μουσική γραμματική του Kohonen στην αναγνώριση κλιμάκων, τυχαίων περιπάτων και ειδικά διεσπαρμένων τυχαίων περιπάτων. Επιπροσθέτως, το CONCERT μοιάζει να συμπεριφέρεται όπως ο άνθρωπος, μιας και όταν του ζητηθεί να παίξει μια κλίμακα έχοντας εκπαιδευτεί μόνο σε κλίμακες η ικανότητα γενίκευσής του πέφτει, και όταν του ζητηθεί να αναπαράγει έναν τυχαίο περίπατο η ικανότητα γενίκευσής του ανεβαίνει.

Από την άλλη, τα ενοχλητικά αποτελέσματα που είχαμε στην δομή AABA μπορούν να προσομοιαστούν με την αντίστοιχη δυσκολία που θα συναντούσε ένας άνθρωπος στην ακουστική κατανόηση αυτής της δομής, όσον αφορά το είδος της φράσης την οποία εξετάσαμε. Φαίνεται δηλαδή, πως θα έπρεπε να είχαμε μνήμη 11 νοτών για να

μπορούσαμε να κατανοήσουμε τις εσωτερικές αλλαγές της δομής αυτής. Παρακάτω θα προσπαθήσουμε να εντοπίσουμε τις δομές ανώτερης τάξης και να τις εντάξουμε στην εκπαίδευση χωρίς να χάσουμε ικανότητα γενίκευσης.

2.7 Εντοπισμός μουσικής δομής μεγαλύτερης τάξης

Απ' ότι παρατηρήθηκε και με το παράδειγμα του CONCERT, η εκπαίδευση του δικτύου με τη μέθοδο της προς τα πίσω ολίσθησης δεν είναι η κατάλληλη μέθοδος για τον εντοπισμό καθολικών δομών. Με τον όρο αναγνώριση καθολικής δομής εννοούμε το πλήθος των στοιχείων που μεσολαβούν από το προηγούμενο στοιχείο μέχρι το προς πρόβλεψη στοιχείο. Αν δηλαδή, το στοιχείο που πρόκειται να ακολουθήσει καθορίζεται από το αμέσως προηγούμενο, τότε έχουμε αναγνώριση τοπικής δομής. Αν αρχίσουμε να παρεμβάλλουμε στοιχεία μεταξύ των δύο αυτών στοιχείων, τότε αρχίζουμε να χρειαζόμαστε αναγνώριση πιο καθολικών δομών [45, 46, 48, 58]. Υπάρχουν όμως θεωρητικά επιχειρήματα πως τα αναδραστικά νευρωνικά δίκτυα έχουν ένα όριο στο πόσο πίσω στο χρόνο μπορούν να μαθαίνουν [3].

Ένα μουσικό κομμάτι θα μπορούσε να περιγραφεί ως μία συλλογή από εναλλασσόμενες φράσεις. Για την αναγνώριση των φράσεων μία προσέγγιση με πίνακες μετάβασης, μια προσέγγιση δηλαδή που εντοπίζει την τοπική δομή, θα μπορούσε να λειτουργήσει, η αναγνώριση όμως των εναλλαγών των φράσεων απαιτεί ένα μέσο αναγνώρισης καθολικότερης δομής.

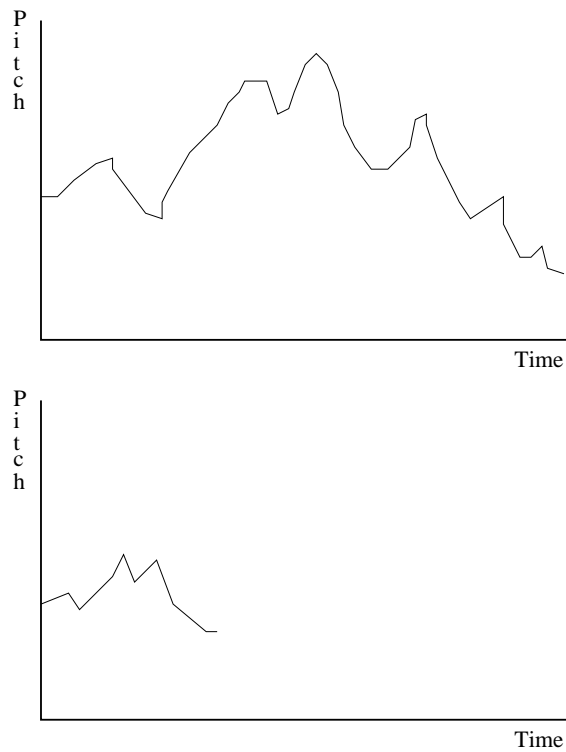
Το παραπάνω πρόβλημα έχει προσεγγιστεί με την κατασκευή μιας *μειωμένης περιγραφής* [46, 48, 49, 55, 58, 59] της σειράς των νοτών η οποία κάνει πιο προφανώς αναγνωρίσιμες τις καθολικές δομές. Στην περίπτωση της δομής AABA, η μειωμένη περιγραφή θα ήταν σε θέση να αναγνωρίσει την δομή A και να την περιγράψει απλά σαν A.

Για να επιτευχθεί η μειωμένη περιγραφή, χρησιμοποιούμε κάτι που θα μπορούσε να περιγραφεί μεταφορικά ως *ευρυγώνιος φακός* για να δούμε το μουσικό κομμάτι. Έτσι μπορούμε να επιτύχουμε κάτι που θα μπορούσαμε να το φανταστούμε σαν να ακούμε ένα τραγούδι παιγμένο σε διπλάσια ταχύτητα. Δεν θα ήμασταν σε θέση, ακούγοντας σε αυτή την ταχύτητα, να καταλάβουμε ξεχωριστά τις νότες μιας και παρουσιάζονται πιο κοντά η μία στην άλλη, μπορούμε όμως να αντιληφθούμε τις αλλαγές στο ύψος που επέρχονται από την αλλαγή των φράσεων. Όπως φαίνεται και στο σχήμα 2.6, το διάγραμμα των νοτών συναρτήσει του χρόνου γίνεται πιο λείο και συμπιεσμένο. Συμβαίνει κάτι ανάλογο με την εφαρμογή φίλτρου που αφήνει να περάσουν οι χαμηλές συχνότητες σε ένα σήμα.

Σύμφωνα με τη παραπάνω οπτική, θα χρησιμοποιήσουμε τις ίδιες μονάδες πλαισίου, μόνο που θα τις αφήσουμε να λειτουργούν με διαφορετικές σταθερές χρόνου. Με μία απλή τροποποίηση του κανόνα ενεργοποίησης των μονάδων πλαισίου έχουμε την παρακάτω εξίσωση ενεργοποίησης

$$c_i(n) = t_i c_i(n-1) + (1-t_i) s \left[\sum_j w_{ij} x_j(n) + \sum_j v_{ij} c_j(n-1) \right]$$

όπου η κάθε μονάδα περιεχομένου i έχει μία αντίστοιχη σταθερά t_i η οποία παίρνει τιμές από το 0 ως το 1 και καθορίζει τον βαθμό στον οποίον η δραστηριότητά του αλλάζει. Με $t = 0$ η παραπάνω συνάρτηση μετατρέπεται στη συνάρτηση ενεργοποίησης



Σχήμα 2.6: Μοντελοποίηση για την αναπαράσταση της διάρκειας.

που είχαμε μέχρι τώρα. Όσο το t μεγαλώνει τόσο πιο 'βραδυκίνητη' γίνεται μονάδα, μένοντας περισσότερο στην τιμή που είχε στην προηγούμενη επανάληψη. Στην ακραία περίπτωση που $t = 1$ η επόμενη τιμή είναι ακριβώς ίδια με την προηγούμενη και η παραπάνω συνάρτηση είναι η σταθερή.

Η συμπάγεια που προδίδει αυτή η σταθερά στην συνάρτηση φαίνεται πιο άμεσα από τη διαδικασία εκπαίδευσης. Αυτό θα φανεί αν δούμε τη σχέση που έχει η παράγωγος του σφάλματος κάθε μονάδας στο επίπεδο ξεδιπλώματος n , $\frac{\partial E}{\partial c_i(n)}$, και το λάθος που ολισθαίνει προς τα πίσω, σε κάποια μονάδα πριεχομένου στο προηγούμενο επίπεδο, $n - 1$, έχουμε πως

$$\frac{\partial E}{\partial c_i(n-1)} = \frac{\partial E}{\partial c_i(n)} \frac{\partial}{\partial c_i(n-1)} [t_i c_i(n-1) + (1-t_i)SE] = t_i \frac{\partial E}{\partial c_i(n)} + (1-t_i)SE$$

, όπου SE το σφάλμα της μονάδας που θα είχαμε χωρίς τη μειωμένη περιγραφή, που σημαίνει ότι όσο μεγαλύτερο είναι το t_i τόσο περισσότερο σφάλμα από το επίπεδο n στο επίπεδο $n - 1$ μεταφέρεται στη μονάδα i . Αυτό μας λέει πως το ολισθαίνον προς τα πίσω σφάλμα μπορεί να επικοινωνήσει με πιο μακρινές παρελθοντικές στιγμές, πράγμα που κάνει δυνατή τη κατανόηση πιο καθολικών δομών στη σειρά εισόδου.

Ακολουθούν σχόλια αναφορικά με αυτή την προσέγγιση.

- Στο σχήμα 2.6 υπάρχει μια σχετική ανακρίβεια γιατί παρουσιάζει το φιλτράρισμα να γίνεται στη σειρά εισόδου πριν μπει στο δίκτυο, ενώ στη πραγματικότητα συμβαίνει το φιλτράρισμα της σειράς αφού έχει μπει στο δίκτυο.
- Οι σταθερές του χρόνου που χρησιμοποιούνται έχουν χρησιμοποιηθεί και στο

παρελθόν με διαφοροποιήσεις στην αρχιτεκτονική των νευρωνικών δικτύων με αναδραστικά δίκτυα [9], εμπρόσθιας τροφοδότησης [41] και συνεχούς χρόνου [52,53] δίκτυα, δεν έχουν όμως λειτουργήσει έτσι ώστε να ελέγχουν την χρονική ανταπόκριση ξεχωριστών μονάδων.

- Υπάρχουν πολλές τιμές του t_i που θα μπορούσαν να χρησιμοποιηθούν για τον ισορροπημένο εντοπισμό τοπικής και καθολικής δομής. Η τιμή που θα διατηρεί την ισορροπία αυτή θα μπορούσε να βρεθεί χρησιμοποιώντας τη μέθοδο της προς τα πίσω ολίσθησης του σφάλματος, υπάρχει όμως ο κίνδυνος για εγκλωβισμό της μεθόδου σε τοπικά ελάχιστα.
- Αυτή η προσέγγιση εξειδικεύει κάθε μονάδα πλαισίου διαφορετική χρονική κλίμακα. Θα μπορούσαμε να πούμε λοιπόν, πως το CONCERT μπορεί να ειδικευτεί σε πολλές χρονικές κλίμακες.
- Ο συγκεκριμένος τρόπος για μείωση της περιγραφής θα μπορούσε να χαρακτηριστεί πολύ απλός. Στην τελευταία ενότητα αυτής της εργασίας αναφερόμαστε σε μια μέθοδο μείωσης της περιγραφής που θα χαρακτηρίζει τεμάχια των γεγονότων εισόδου τα οποία χαρακτηρίζονται από νοηματική συνοχή.

2.8 Επαναπροσέγγιση των φράσεων AABA

Στο πείραμα που έγινε πριν για την εκπαίδευση στις φράσεις AABA, χρησιμοποιήθηκαν 35 μονάδες περιεχομένου χωρίς μείωση της περιγραφής, δηλαδή με $t = 0$. Τώρα θα χρησιμοποιήσουμε 30 μονάδες περιεχομένου, 25 από τις οποίες έχουν $t = 0$ και οι υπόλοιπες 5 έχουν $t = 0.8$. Ο πίνακας 2.8 δείχνει την επίδοση στη γενίκευση των δύο πειραμάτων αναφορικά με την τοπική και την ολική δομή. Η επίδοση για την κατανόηση καθολικής δομής είναι σημαντικά καλύτερη με τη μειωμένη περιγραφή και η κατανόηση τοπικής δομής ήταν οριακά χειρότερη.

Σε πειράματα που δοκιμάστηκαν τιμές του t μεταξύ 0.7 και του 0.95 τα αποτελέσματα ήταν ποιοτικά όμοια, ακόμα και στα πειράματα που οι φράσεις A και B ήταν τυχαίοι μουσικοί περίπατοι στο κλειδί της C μείζονος. Αν και η κατανόηση της καθολικής δομής βελτιώνεται, ποτέ δεν ξεπερνά την κατανόηση της τοπικής, πράγμα που μας δείχνει πως οι ικανότητες του CONCERT μοιάζουν πολύ με τις ικανότητες του ανθρώπου σε αυτό το απλό ζήτημα.

είδος δομής	αρχικό δίκτυο	δίκτυο με μειωμένη περιγραφή
τοπική	97.3%	96.7%
καθολική	58.4%	75.6%

Πίνακας 2.8: Επίδοσεις των δικτύων με και χωρίς μειωμένη περιγραφή στις φράσεις AABA.

2.9 Πειράματα μεγαλύτερης κλίμακας

Στις επόμενες ενότητες που ακολουθούν θα περιγράψουμε την ανταπόκριση του CONCERT με σύνολο εκπαίδευσης πραγματικά κομμάτια μουσικής, χρησιμοποιώντας και την τεχνική της μείωσης της περιγραφής.

2.9.1 Σύνθεση μελωδίας στη μορφή συνθέσεων του Bach

Οι μελωδικές γραμμές δέκα απλών κομματιών του J. S. Bach χρησιμοποιήθηκαν για την εκπαίδευση του CONCERT με σχετικά μεγάλο αριθμό νοτών το κάθε ένα. Τα κομμάτια του Bach βρίθουν κλιμάκων και διαφορετικών μουσικών μορφών, γεγονός που τα κάνει πολύ δύσκολα παραδείγματα για το νευρωνικό μας δίκτυο. Μπορούμε να πούμε πως το μόνο κοινό που έχουν τα μουσικά κομμάτια του Bach είναι ο συνθέτης τους. Τα κομμάτια αυτά έχουν πολλές μουσικές μελωδικές γραμμές από τις οποίες χρησιμοποιούμε την κεντρική μελωδία. Το σημαντικό είναι πως σε έναν ανεκπαιδευτο μουσικά άνθρωπο οι μελωδίες αυτές από μόνες τους, χωρίς τη συνοδεία συγχορδιών και άλλων φωνών, ακούγονται ευχάριστες και συνεκτικές.

Για να καταλαβαίνει το CONCERT πότε ένα κομμάτι τελειώνει, το τέλος κάθε κομματιού σηματοδοτείται με τρεις συνεχόμενες παύσεις. Επιπλέον, κάθε κομμάτι που είναι σε μείζων κλειδί μετατράπηκε σε C μείζονα, ενώ τα κομμάτια που είναι σε ελάσσων κλειδί μετατράπηκαν στο κλειδί σε A ελάσσονα. Με μια διαφορετική αναπαράσταση αυτή η μετατροπή μπορούσε να αποφευχθεί.

Δύο σταθερές μονάδες έχουν συμπεριληφθεί στην είσοδο του CONCERT για αυτά τα πειράματα. Η μία καθορίζει το αν το κομμάτι είναι σε κλειδί μείζων ή ελάσσων και η άλλη για το αν το κομμάτι είναι σε μέτρο $\frac{4}{4}$, $\frac{3}{4}$, $\frac{2}{4}$. Οι αξίες αυτών των μονάδων μένουν αμετάβλητες κατά τη διάρκεια του κομματιού. Μία επιπλέον μονάδα προστέθηκε για να κρατάει το CONCERT εντός μέτρου η οποία είναι ενεργοποιημένη για μία νότα όταν αυτή βρίσκεται στον πρώτο χτύπο του μέτρου. Αναφορικά με τις νότες που περνούν το μέτρο στο οποίο ξεκίνησαν, τις χειριζόμαστε σαν δύο διαφορετικές νότες. Αυτό για να διασφαλίσουμε το ότι κάθε πρώτος χτύπος του κάθε μέτρου γίνεται αντιληπτός. Για να διατηρούμε πληροφορίες για το πότε έχουμε συζευγμένη νότα, προστέθηκε ένα επιπλέον στοιχείο για να ενημερώνει αν η παρούσα νότα είναι συζευγμένη με τη προηγούμενη.

Η εκπαίδευση του CONCERT περιλαμβάνει την εκπαίδευση σε πρόβλεψη συνόλου 1260 νοτών που αποτελούν τις νότες 10 κομματιών του J. S. Bach μαζί με τις παύσεις που τα χωρίζουν. Η εκπαίδευση έγινε με 40 κρυφές μονάδες, 35 από τις οποίες είχαν $t = 0$ και 5 από αυτές $t = .8$ για 3000 περάσματα με το σύνολο εκπαίδευσης. Ο βαθμός μάθησης μειωνόταν σταδιακά από .0004 σε .0002. Με την ολοκλήρωση της εκπαίδευσης το CONCERT μπορούσε να προβλέψει το 95% από τις τονικότητες και το 95% από τις διάρκειες σωστά. Σε προσπάθειες που έγιναν για εκπαίδευση του CONCERT με μεγαλύτερη μείωση της περιγραφής, δηλαδή με περισσότερες μονάδες περιεχομένου με μεγαλύτερες τιμές του t , οι επιδόσεις ήταν χειρότερες.

Για να λειτουργήσουμε το CONCERT σε μορφή σύνθεσης το τροφοδοτούμε με τις τέσσερις πρώτες νότες ενός κομματιού από το σύνολο εκπαίδευσης και το αφήνουμε να προβλέψει τις επόμενες. Οι επιλογή των νοτών και της διάρκειας τους γίνεται ανεξάρτητα, ανάλογα με την πιθανότητα που ακολουθεί, με έναν επιπρόσθετο περιορισμό για τη διάρκεια σύμφωνα με τον οποίον μία διάρκεια θεωρείται άκυρη αν έχει περάσει τα όρια του μέτρου (πράγμα το οποίο θα ήταν επιτρεπτό μόνο αν ήταν ενεργή η ένδειξη της συζευγμένης νότας). Μόλις το CONCERT αρχίζει να συνθέτει, γρήγορα το αποτέλεσμα αρχίζει να διαφέρει από το παράδειγμα εκπαίδευσης από το οποίο ξεκίνησε, δίνοντας όμως μερικά σημεία που μοιάζουν με σημεία του κομματιού αυτού. Το κομμάτι τελειώνει μόλις παρατίθεται μία σειρά από παύσεις.

Μερικές συνθέσεις δημιουργήθηκαν και με δύο πίνακες μετάβασης τρίτης τάξης,

έναν για την τονικότητα και έναν για τη διάρκεια. Αν και ο όγκος των πληροφοριών που χρειαζότανε για την κατασκευή αυτών των πινάκων ήταν τεράστιος, το αποτέλεσμα όταν κρίθηκε από 12 μουσικά ανεκπαιδευτους ακροατές θεωρήθηκε χειρότερο από αυτό που παρήγαγε το CONCERT. Οι παρουσιάσεις των αποτελεσμάτων έγιναν εναλλάξ όμως και τα 12 άτομα υποστήριζαν, άλλοι ισχυρά και άλλοι όχι, πως οι συνθέσεις που αντιστοιχούσαν στις συνθέσεις του CONCERT είχαν περισσότερη συνοχή και μεγαλύτερη μελωδικότητα.

2.9.2 Σύνθεση μουσικής στη μορφή συνθέσεων ευρωπαϊκών λαϊκών κομματιών

Σε ένα δεύτερο πείραμα, το CONCERT εκπαιδεύτηκε σε ένα σύνολο από 25 παραδοσιακά ευρωπαϊκά τραγούδια. Τα κομμάτια αυτά ήταν βραχύτερα από τα κομμάτια του Bach, έχοντας 75 νότες μέσο όρο. Όλα τα τραγούδια ήταν στο κλειδί της C μείζονος και το μέτρο τους ήταν $\frac{4}{4}$ και γι'αυτό δεν χρειάστηκαν επιπλέον μονάδες για τον καθορισμό αυτών των γεγονότων, όπως έγινε στο προηγούμενο πείραμα.

Η εκπαίδευση έγινε με 50 κρυφές μονάδες, για 45 από τις οποίες ίσχυε $t = 0$, ενώ για τις υπόλοιπες 5 ίσχυε $t = .8$, για 2000 περάσματα του συνόλου εκπαίδευσης. Με το τέλος της εκπαίδευσης το CONCERT ήταν σε θέση να αναπαράγει σωστά το 93% των τονικοτήτων και το 90% από τις διάρκειες του κάθε κομματιού από το σύνολο εκπαίδευσης. Όταν δε, λειτουργούσε συνθετικά, τα αποτελέσματα ήταν αρκετά πιο ενδιαφέροντα από τα γνήσια!

2.9.3 Αποκλειστική εκπαίδευση σε δομές υψηλής τάξης: το πείραμα του βαλς

Παρότι το CONCERT είναι καλύτερο από τους πίνακες μετάβασης τρίτης τάξης, δύσκολα εντοπίζει δομές υψηλότερης τάξης. Για να γίνει αισθητή η μεγέθυνση του χρόνου στο τελευταίο αυτό πείραμα, λαμβάνουμε υπόψιν και τις συγχορδίες που συνοδεύουν τη μελωδική γραμμή. Το σύνολο εκπαίδευσης σε αυτό το πείραμα αποτελείται από 25 κομμάτια βαλς από γνωστούς συνθέτες, των οποίων η μελωδία απλοποιήθηκε, θεωρώντας και τις συγχορδίες που την συνοδεύουν. Οι συγχορδίες παραπέμπουν σε μεγαλύτερη χρονική κλίμακα αφού η μέση περίοδος αλλαγής νοτών είναι 1.4 χτύποι, ενώ η αντίστοιχη των συγχορδιών είναι 5.9. Έτσι λοιπόν, δίνεται η ελπίδα πως το CONCERT θα μπορούσε να εντοπίσει μεγαλύτερης τάξης δομή αναγνωρίζοντας τις συγχορδίες.

Στο συγκεκριμένο πείραμα οι συγχορδίες που συνόδευαν τη μελωδία, δεν είχαν αποκλειστικά δικό τους μέτρο διάρκειας αλλά η διάρκειά τους χαρακτηριζόταν από το άθροισμα των διαρκειών των νοτών πάνω από τις οποίες ήταν ενεργοποιημένες. Το πείραμα αυτό, δυστυχώς, δε μας έδειξε πως η καθολική δομή δεν μπορούσε να γίνει κατανοητή.

2.10 Συζήτηση για την συνολική απόδοση του CONCERT

Αν και το CONCERT δουλεύει καλά απλές, τεχνητές δομές, στην πραγματική μουσική δεν ανταποκρίνεται πειστικά. Δύσκολα κάποιος ακροατής θα μπορούσε

να ξεγελαστεί και να θεωρήσει τις συνθέσεις του ανθρώπινες. Αυτό γιατί αν και σε διάφορα μέρη υπάρχει τοπική συνοχή, η έλλειψη καθολικής δομής στα κομμάτια που συνθέτει δημιουργεί κουραστικά ακούσματα, μιας και δεν υπάρχει κεντρικό θέμα, οι μεταβάσεις με μετατροπές γίνονται πολύ κάθετα και άλλα.

Η αρχιτεκτονική και ο τρόπος εκπαίδευσης του CONCERT φαίνεται πως δεν του δίνουν τη δυνατότητα να κατανοεί καθολικές δομές στη μουσική και να αναπαράγει ολοκληρωμένα μουσικά κομμάτια. Αυτό θα μπορούσε κάποιος να το θεωρήσει λογικό, μιας και κατά αντιστοιχία η εκπαίδευση σύνθεσης μουσικής νότας προς νότα είναι σαν την διδασκαλία για την εκμάθηση γραφής και ομιλίας μιας γλώσσας με ανάλυση γράμμα προς γράμμα. Άρα οι έρευνες που έγιναν πάνω στο CONCERT δίνουν πεσιμιστικά μηνύματα για την σύνθεση μουσική με την προσέγγιση νότας προς νότα, ακόμη και αν χρησιμοποιηθούν πανίσχυροι αλγόριθμοι για την εκπαίδευση.

Παρακάτω αναφέρουμε μερικές τεχνικές που θα μπορούσαν να βοηθήσουν ένα νευρωνικό δίκτυο να καταλάβει την καθολική δομή ενός μουσικού κομματιού.

- *Πολλαπλά αναδραστικά κρυφά στρώματα.* Το CONCERT θα μπορούσε να βελτιωθεί με τη χρήση πολλαπλών κρυφών στρωμάτων στη αναδραστικό μηχανισμό του, έτσι ώστε να υπάρχει μεγαλύτερη μη γραμμική πολυπλοκότητα στην εκπαίδευση αλλά και στα κριτήρια επιλογής της επόμενης νότας [12].
- *Εναλλακτικές προσεγγίσεις στην μνήμη.* Ένας τρόπος που θα μπορούσε το CONCERT να θυμάται τα γεγονότα του παρελθόντος είναι η χρήση μιας προσωρινής μνήμης [48] στις μονάδες εισόδου ή/και στις μονάδες περιεχομένου έτσι ώστε να υπάρχει μνήμη για τις χρονικά κοντινές περιοχές χωρίς να διασπρέβλονται η πληροφορία στα αναδραστικά στρώματα, πράγμα που συμβαίνει με τη μείωση της περιγραφής που χρησιμοποιήσαμε.
- *Δίκτυα διαδοχικής ιεραρχίας.* Αφορά νευρωνικά δίκτυα διαδοχικά τοποθετημένα έτσι ώστε να έχουν ιεραρχική δομή. Η ιδέα με αυτά τα δίκτυα είναι τα χαμηλότερα στρώματα της ιεραρχίας να εντοπίζουν τις πιο τοπικές ή λεπτές δομές, ενώ τα υψηλότερα να εντοπίζουν τις πιο καθολικές ή πιο τραχείες δομές [11, 64].
- *Τεμαχισμένη αρχιτεκτονική.* Στην διαδοχική ιεραρχική δομή που περιγράφηκε παραπάνω, οι σχεδιαστές του δικτύου δίνουν εκ των προτέρων την ιεραρχική αποσύνθεση της σειράς εισόδου. Στην τεμαχισμένη αρχιτεκτονική, η σειρά της εισόδου δίνεται τεμαχισμένη και το δίκτυο αποφασίζει από μόνο του τα επίπεδα της ιεραρχίας τα οποία θα χειριστεί. Στην προκειμένη περίπτωση, μία ιδανική εφαρμογή των παραπάνω θα ήταν τα χαμηλότερα ιεραρχικά στρώματα να έδιναν τεμάχια με νόημα στα παραπάνω ιεραρχικά στρώματα προς επεξεργασία [48, 59].
- *Ευθεία αναπαράσταση της δομής.* Όπως στο τελευταίο πείραμα χρησιμοποιήθηκε η εκπαίδευση του δικτύου πάνω στις συγχορδίες, έτσι ώστε να αντιληφθεί την τραχύτερη δομή στην οποία αυτές υπακούν, θα μπορούσαμε να αναγκάσουμε το δίκτυο να εκπαιδευτεί με βάση τις μουσικές φράσεις, τα μουσικά θέματα, τις αναστροφές ή το τονικό κλειδί εντός του κομματιού.
- *Κλιμακούμενη εκπαίδευση.* Έχοντας σαν δεδομένο ότι το CONCERT μπορεί να χειριστεί την τοπική δομή, κάποιος θα μπορούσε να το εκπαιδεύσει κλιμακωτά, ξεκινώντας από μεγάλη μείωση της περιγραφής και κατά τη διάρκεια της

εκπαίδευσης να μειώνεται η μείωση της περιγραφής. Με αυτόν τον τρόπο το CONCERT θα μπορούσε να λαμβάνει υπόψιν του την καθολική δομή από την αρχή της εκπαίδευσής του.

- *Αναπαράσταση μουσικών στοιχείων στο περιεχόμενο.* Στην εργασία αυτή, η αναπαράσταση της τονικότητας, της διάρκειας και των συγχορδιών είναι ανεξάρτητη από το περιεχόμενο στο οποίο είναι εμβαπτισμένες. Τα ψυχοακουστικά κριτήρια με τα οποία επιλέχθηκαν δεν αφορούν το περιεχόμενο αυτό καθαυτό, αλλά την πολύ μικρή γειτονιά της κωδικοποιούμενης νότας. Τα ψυχοακουστικά κριτήρια δηλαδή, εφαρμόστηκαν μετά από εξέταση στις αντιδράσεις ανθρώπων σε ζευγάρια νοτών και όχι σε μεγαλύτερα τμήματα. Αυτό είναι όμως λάθος, γιατί στη μουσική, όπως και σε άλλα γνωστικά πεδία, το μέχρι κάποια στιγμή περιεχόμενο και η αναμονή για το υπόλοιπο επηρεάζουν την αντίληψη του στοιχείων [28, 29, 32, 33]. Μια πιο αληθοφανής αρχιτεκτονική θα μπορούσε να επιτρέπει αλληλεπιδράσεις μεταξύ αποκωδικοποίησης και της διαδικασίας, η οποία θα κωδικοποιούσε την αναμονή της συνέχειας του περιεχομένου. Όμοια κωδικοποίηση των διαρκειών και των συγχορδιών θα μπορούσαν να εντοπίζουν καθολικότερες δομές [8].

Αναγνώριση Μουσικού Συνθέτη με χρήση Τεχνητών Νευρωνικών Δικτύων

3.1 Εισαγωγή

Η αναγνώριση του συνθέτη ενός μουσικού κομματιού είναι εδώ και πολλά χρόνια αντικείμενο μελέτης τόσο της μουσικολογίας όσο και των μαθηματικών. Με τη μελέτη πάνω στην αναγνώριση του συνθέτη ενός μουσικού κομματιού δεν επιτυγχάνεται μόνο η ταυτοποίηση του προσώπου που συνέθεσε κάποια κομμάτια. Κυρίως επιχειρείται μια βαθύτερη προσέγγιση του τι τελικά σημαίνει μουσική. Πως κάποιες συχνότητες ήχου οι οποίες κάνουν τα τύμπανα στα ανθρώπινα αυτιά να πάλλονται, έχουν τόσο μεγάλη επίδραση στα συναισθήματα.

Η αναζήτηση αυτή γίνεται έμμεσα καθώς αντικείμενα του μουσικού φορμαλισμού (τονικότητες, διάρκειες, παύσεις) δεν αναλύονται ως έχουν, παρά γίνεται αντιπαράθεση του τρόπου χρήσης τους από συνθέτη σε συνθέτη. Κάποιοι μουσικοί συνθέτες ανήκουν στην εποχή τους, κάποιοι άλλοι όμως στην αιωνιότητα. Τι μπορεί να είναι αυτό που ξεχωρίζει τους μεν από τους δε συνθέτες; Τι ξεχωρίζει σπουδαίους μουσικούς συνθέτες μεταξύ τους;

Έχουν γίνει διάφορες στατιστικές προσεγγίσεις σε παρτιτούρες και σε κυματομορφές μουσικών κομματιών [5, 6, 23]. Αυτές οι προσεγγίσεις δίνουν έναν εναλλακτικό τρόπο για την κατηγοριοποίηση των κομματιών ανάλογα με την εποχή που συντέθηκαν ή τον συνθέτη που τα συνέθεσε εξετάζοντας κάποια στατιστικά στοιχεία.

Από την άλλη, κάποιοι ειδικός μουσικολόγος, κάποιος καταρτισμένος μουσικός ή ακόμα και ένα μαθητής μουσικής έχουν αντιμετωπίσει πολλές φορές την πρόκληση να αναγνωρίσουν τις καταβολές ενός μουσικού κομματιού. Όταν κάποιος μουσικός για παράδειγμα ερμηνεύει ένα κομμάτι κάποιου συνθέτη, προσπαθεί με την ερμηνεία του να αποδώσει την αίσθηση που του υπαγορεύει ο συνθέτης αυτός.

Ένας μουσικός είναι εκπαιδευμένος να εντοπίζει μορφώματα σε διαφορετικά επίπεδα πάνω σε μια παρτιτούρα στην προσπάθειά του να αναγνωρίσει το συνθέτη ενός κομματιού. Είναι γνωστό ότι ο ανθρώπινος εγκέφαλος είναι ένα μεγάλο νευρωνικό δίκτυο που συνδυάζει ερεθίσματα από διαφορετικές αισθήσεις έτσι ώστε να παίρνει αποφάσεις με μεγάλη γενίκευση. Αυτή η ικανότητα κληροδοτήθηκε από τους ανθρώπους στους υπολογιστές μέσω των τεχνητών νευρωνικών δικτύων.

Δε θα είχε νόημα να βάλουμε τα τεχνητά νευρωνικά δίκτυα να συναγωνιστούν με αυτά του ανθρώπινου εγκέφαλου χρησιμοποιώντας την ίδια μοντελοποίηση για την υλοποίηση του συναγωνισμού αυτού. Το ποιος θα ήταν ο νικητής μας το μαρτυρά το πλήθος των νευρώνων που υπάρχουν στον ανθρώπινο εγκέφαλο και η πλούσια ποικιλία στις συνάψεις που δημιουργούν. Δε θα είχε λοιπόν νόημα να μοντελοποιή-

σουμε τον τρόπο που ένα μουσικός εξετάζει την πατρότητα ενός μουσικού κομματιού και να περάσουμε αυτά τα μοντέλα στα νευρωνικά δίκτυα, γιατί ο ανθρώπινος εγκέφαλος μπορεί σίγουρα να τα διαχειριστεί καλύτερα. Θα είχε όμως νόημα να δούμε πως ένα τεχνητό νευρωνικό δίκτυο, απελευθερωμένο από τις αισθήσεις που οδηγούν τους ανθρώπους σε αποφάσεις, διαχειρίζεται στοιχεία που ο ανθρώπινος εγκέφαλος δεν μπορεί να διαχειριστεί. Έτσι θα μπορούσαμε να έχουμε μια συνεργασία της τεχνητής με την ανθρώπινη νοημοσύνη για καλύτερα αποτελέσματα.

3.2 Εξόρυξη πληροφοριών κειμένου μέσω παρτιτούρας

Η παρτιτούρα ενός μουσικού κομματιού είναι ένα κείμενο μιας συγκεκριμένης μορφής που μπορεί να διαβαστεί από κάποιο μουσικά καταρτισμένο. Ο τρόπος που ένας υπολογιστής μπορεί να διαβάσει και να διαχειριστεί μια παρτιτούρα είναι μέσω της γλώσσας των *MIDI* αρχείων.

3.2.1 Τι είναι τα MIDI αρχεία

Τα αρχεία MIDI (*Musical Instrument Digital Interface*) είναι ένας τρόπος μοντελοποίησης της μουσικής από τους υπολογιστές. Ο χρόνος τοποθέτησης μιας νότας, η διάρκεια, η ένταση, η τονικότητα και το μουσικό όργανο που την αναπαράγει δίνονται στον υπολογιστή σαν εντολές που ακολουθούν ένα συγκεκριμένο πρωτόκολλο. Ο υπολογιστής αποκωδικοποιεί αυτές τις εντολές και παράγει τον αντίστοιχο ήχο χρησιμοποιώντας ένα εσωτερικό κύκλωμα (*synthesiser*) της κάρτας ήχου του.

Η ποιότητα ήχου των αρχείων MIDI είναι φτωχή, όμως το αντάλλαγμα είναι το πολύ μικρό τους μέγεθος. Αυτό τα αρχεία είναι εντολές, γραμμές κώδικα που εύκολα χωράνε μέσα σε ένα αρχείο κειμένου, και όχι κυματομορφές που η αποκωδικοποίησή τους απαιτεί μεγάλες ποσότητες πληροφορίας. Επιπροσθέτως, τα αρχεία MIDI μας είναι χρήσιμα γιατί περιέχουν όλη την πληροφορία που έχει μια παρτιτούρα. Υπάρχουν επίσης προγράμματα που μετατρέπουν μια παρτιτούρα σε MIDI αρχείο και αντίστροφα (*finale*, *cakewalk*, *rosegarden* κ.α.).

3.2.2 Πως χρησιμοποιούμε τα αρχεία MIDI

Τα μουσικά κομμάτια που χρησιμοποιήσαμε τα πήραμε σε μορφή MIDI από τις παρακάτω ηλεκτρονικές διευθύνσεις

<http://www.piano-midi.de/>
<http://www.classicalmusicmidipage.com/>

όπου και διανέμονται δωρεάν. Για να γίνει η επιθυμητή μοντελοποίηση έπρεπε τα αρχεία αυτά να μετατραπούν σε αναγνώσιμη μορφή κειμένου. Αυτό επιτεύχθηκε με το πρόγραμμα *midi2abc* που κάνει αυτή τη μετατροπή μέσω ενός συγκεκριμένου πρωτοκόλλου, η αποκωδικοποίηση του οποίου μας οδήγησε στην επιθυμητή μοντελοποίηση που θα περιγράψουμε παρακάτω. Τα αρχεία κειμένου που τελικά αποκομίσαμε από τη μέχρι τώρα διαδικασία τα διαχειριστήκαμε με τη γλώσσα προγραμματισμού *PERL*.

3.2.3 Ουσιώδη στοιχεία μιας παρτιτούρας

Το ερώτημα που τίθεται σε αυτήν την υποενότητα είναι το ποια στοιχεία του κειμένου, το οποίο προέρχεται από μια παρτιτούρα, που τελικά έχουμε στα χέρια μας χρήζουν προσοχής και βρῖθουν στατιστικά χρήσιμων στοιχείων. Όπως εξηγήσαμε παραπάνω, δε θα προσπαθήσουμε να βάλουμε ένα τεχνητό νευρωνικό δίκτυο να συναγωνιστεί έναν μουσικό, παρά θα δούμε αν μπορεί να βρει *καυονικότητες* μέσα σε στοιχεία που το ανθρώπινο μάτι κρίνει στατιστικά ασήμαντα.

Παραπάνω αναφέραμε πως υπάρχει μεγάλος πλούτος πληροφοριών που μπορούμε να εξαγάγουμε από μια παρτιτούρα. Όταν ζητείται από ένα μουσικό να αναγνωρίσει το συνθέτη ενός κομματιού εξετάζει πολλές ομάδες πληροφοριών ταυτόχρονα για να αποφασίσει: την πυκνότητα των νοτών (διάρκειες), διαστήματα, αλλοιώσεις εντός της κλίμακας, χρήση της πολυφωνίας (διαχείριση των φωνών), εντοπισμός μοτίβων και πολλά άλλα. Παίρνοντας σαν δεδομένο ότι αυτή ή αυτός ο μουσικός έχει εξετάσει μια πληθώρα από παρτιτούρες και έχει υπόψη του και την ηχητική προβολή τους (τη μουσική που περιγράφουν), φαίνεται να έχει πλεονέκτημα έναντι ενός υπολογιστή που διαχειρίζεται μόνο αριθμούς.

Η στατιστική είναι η γέφυρα μεταξύ παρατηρήσεων και μαθηματικών. Η εφαρμογή της στατιστικής στη μουσική είναι λίγο πολύ μια προσπάθεια να δημιουργήσουμε μαθηματικά μοντέλα για τον τρόπο που ένας ειδικός μουσικός εξετάζει μια παρτιτούρα, δηλαδή η θεώρηση των στοιχείων που εξετάζει ο μουσικός ως στατιστικά σημαντικά.

Προς το παρόν τα ποσοστά από νότες που ένα μουσικό κομμάτι περιέχει δε θεωρείται στατιστικά σημαντικό στοιχείο. Η χρήση των ποσοστών νοτών έχει χρησιμοποιηθεί μέχρι τώρα για την κατηγοριοποίηση κομματιών σε περιόδους [7, 23]. Ένας ειδικός μουσικός δεν μπορεί να κρίνει αν ένα κομμάτι ανήκει για παράδειγμα στον Chopin ή όχι κοιτώντας ένα χαρτί με ποσοστά νοτών. Φαίνεται πως οι άνθρωποι μπορούν καλύτερα να κρίνουν χρησιμοποιώντας έναν συνδυασμό αισθήσεων. Οι ελάχιστες αλλαγές που μπορεί να συμβαίνουν στα ποσοστά των νοτών μεταξύ ομοιάζοντων συνθετών δίνουν ελάχιστα στοιχεία σε κάποιον για να αποφασίσει για την πατρότητα ενός κομματιού σε σχέση με τα στοιχεία που του δίνει μια γρήγορη ματιά σε παρτιτούρες.

Ένα πρόημο συμπέρασμα φαίνεται πως είναι ότι τα σημαντικά στοιχεία που θα μπορούσαν να εξαχθούν από μια παρτιτούρα ποικίλουν ανάλογα με τη μηχανή που τα διαχειρίζεται. Το ανθρώπινο μυαλό που δέχεται ερεθίσματα από πολλές αισθήσεις μπορεί να αποφασίσει κρίνοντας βάσει πολλών στοιχείων που έχουν σημαντικές στατιστικές διακυμάνσεις, ενώ ένα τεχνητό νευρωνικό δίκτυο μπορεί να διαχειριστεί πιο *εκλεπτυσμένες* δομικές διαφορές σε μία συγκεκριμένη περιοχή γεγονότων. Ας δούμε αν οι παραπάνω παρατηρήσεις ισχύουν στην αναγνώριση μουσικού συνθέτη.

3.3 Μια πρώτη προσέγγιση στην αναγνώριση συνθέτη από ΤΝΔ

3.3.1 Μια απλή προσέγγιση

Σκοπός μας είναι να εξετάσουμε αν ένα ΤΝΔ μπορεί να αναγνωρίσει επιτυχώς το συνθέτη ενός κομματιού παίρνοντας πληροφορίες μόνο για τα ποσοστά νοτών που χρησιμοποιούνται σε αυτό το κομμάτι. Η ιδέα πως ένας συνθέτης χρησιμοποιεί

58 Αναγνώριση Μουσικού Συνθέτη με χρήση Τεχνητών Νευρωνικών Δικτύων

συγκεκριμένες ομάδες νοτών φαίνεται απλή αν σκεφτούμε την πληθώρα μουσικών θεμάτων που έχουν συντεθεί από συνθέτες σαν τους Chopin, Beethoven, Mozart, Bach και Handel.

3.3.2 Τοποθέτηση των κομματιών σε ισοδύναμα τονικά μέτρα

Για να είμαστε πιο ακριβείς, θέλουμε να ερευνήσουμε πως ένας συνθέτης συμπεριφέρεται σε ένα μείζων ή ένα ελάσσων μουσικό θέμα. Για να γίνουμε πιο κατανοητοί, θα θεωρήσουμε μια σύνθεση όχι σαν ένα σύνολο νοτών αλλά σαν ένα σύνολο βαθμίδων σε μία μείζονα κλίμακα.

Για παράδειγμα, ας θεωρήσουμε το παρακάτω θέμα στην C μείζονα κλίμακα :

C, E, G, E, C, E, G, Bb

Διαχειριζόμενοι το παραπάνω θέμα σαν βήματα στη C μείζονα κλίμακα θα ήταν ως εξής:

1, 3, 5, 3, 1, 3, 5, b7

όπου το 1 είναι η πρώτη βαθμίδα της C μείζονος κλίμακας. Θεωρώντας τώρα το παρακάτω μουσικό θέμα στη E ύφεση μείζονα κλίμακα :

Eb, G, Bb, G, Eb, G, Bb, Db

θα το διαχειριζόμασταν ως εξής:

1, 3, 5, 3, 1, 3, 5, b7

σαν βήματα της E ύφεση κλίμακας. Ενώ τα δύο παραπάνω μουσικά θέματα (όσο τετριμμένα και αν είναι) αποτελούνται από διαφορετικές νότες, με μια αναπαράσταση σαν αυτή που προτείνεται γίνονται ίδια.

Ένα θέμα που βρίσκεται σε μία ελάσσονα κλίμακα μπορούμε να κάνουμε ακριβώς το ίδιο με παραπάνω μόνο που θεωρούμε ως τονικό κέντρο την έκτη βαθμίδα της αντίστοιχης μείζονος κλίμακας. Για παράδειγμα το παρακάτω μουσικό θέμα στη σολ ελάσσονα κλίμακα :

G, D, Bb, G, F, G

θα το διαχειριζόμασταν ως εξής:

6, 3, 1, 6, 5, 6

όπου 1 είναι η πρώτη βαθμίδα της σι ύφεση μείζονος κλίμακας, που είναι η αντίστοιχη μείζονα της σολ ελάσσονα κλίμακας.

Ένα μουσικό κομμάτι είναι σαν ένα ταξίδι, ξεκινά από κάποιο μέρος και ακολουθώντας μια πορεία φτάνει σε κάποιο μέρος (κάποιες φορές εκεί απ' όπου ξεκίνησε). Κάθε συνθέτης δίνει εναλλακτικές πορείες και σημεία άφιξης ακόμα και αν τα σημεία εκκίνησης είναι παρόμοια με κάποιου άλλου συνθέτη. Αυτό σημαίνει πως κάθε μονοπάτι που ένα μουσικό κομμάτι δημιουργεί αφήνει τα δικά του ίχνη τα οποία θα προσπαθήσουμε να εντοπίσουμε μετρώντας βαθμίδες σε μείζονες κλίμακες. Θέτουμε *κοινό σημείο εκκίνησης* για όλα τα κομμάτια που έχουν ως κεντρικό θέμα ένα θέμα της μείζονος κλίμακας και άλλο ένα γι' αυτά που έχουν θέμα στην ελάσσονα κλίμακα.

3.3.3 Τεχνικά ζητήματα

Για να αποφύγουμε την υλοποίηση διαφορετικών προγραμμάτων για κάθε τοπικότητα κεντρικού θέματος που θα συναντήσουμε μετατρέπουμε κάθε κομμάτι με θέμα σε μείζονα κλίμακα σε C μείζονα και σε ελάσσονα στη A ελάσσονα χρησιμοποιώντας το πρόγραμμα *Finale*. Τα κομμάτια αυτά βρέθηκαν σε μορφή αρχείων MIDI όπως περιγράψαμε παραπάνω. Χρησιμοποιώντας το πρόγραμμα *midi2abc* μετατρέψαμε τα παραπάνω αρχεία σε αρχεία κειμένου. Τέλος, υλοποιήθηκαν προγράμματα στη γλώσσα προγραμματισμού *PERL* για οποιαδήποτε τροποποίηση στα αρχεία κειμένου.

Σκοπός μας είναι να μετρήσουμε τα ποσοστά χρήσης συγκεκριμένων βαθμίδων πάνω στη μείζονα κλίμακα, ή ακριβέστερα να μετρήσουμε *πυκνότητες χρήσης* των βαθμίδων αυτών, πληροφορίες για τις οποίες αποκτούμε όπως περιγράφεται παρακάτω.

Υποθέστε πως έχουμε έναν μετρητή για κάθε βαθμίδα. Ο μετρητής της κάθε βαθμίδας είναι μηδέν όταν το κομμάτι ξεκινά. Σαρώνοντας το κομμάτι μέχρι το τέλος, όταν μια βαθμίδα της μείζονος κλίμακας εμφανίζεται τότε ο αντίστοιχος μετρητής αυξάνεται κατά ένα. Αν εμφανιστούν περισσότερες από μια νότες ταυτόχρονα (συγχορδίες ή πολλαπλές φωνές) τότε ο κάθε αντίστοιχος μετρητής αυξάνεται κατά ένα. Ας υποθέσουμε πως σαρώσαμε ένα κομμάτι μέχρι το τέλος και οι μετρητές βαθμίδων της μείζονος κλίμακας έχουν τις τιμές που φαίνονται στον παρακάτω πίνακα :

Βαθμίδες μείζονος κλίμακας	Μετρητής
1	314
#1	8
2	481
#2	26
3	455
4	220
#4	160
5	770
#5	4
6	589
#6	46
7	564

Πίνακας 3.1: Παράδειγμα πλήθους βαθμίδων μείζονος κλίμακας σε τυχαίο κομμάτι.

όπου το 1 δείχνει τη πρώτη βαθμίδα της μείζονος κλίμακας στην οποία συνετέθη του κομμάτι του παραδείγματος.

Η τιμή του μετρητή όμως εξαρτάται και από το μέγεθος του μουσικού κομματιού. Πρέπει λοιπόν με κάποιο τρόπο να *ομαλοποιήσουμε* τους μετρητές έτσι ώστε να μετρούν πυκνότητες χρήσης των βαθμίδων. Η ομαλοποίηση που ακολουθούμε σε αυτήν την εργασία ¹ συμβαίνει διαιρώντας την τιμή του κάθε μετρητή με τη μέγιστη τιμή όλων των μετρητών. Έτσι δημιουργούνται τιμές πυκνοτήτων στους μετρητές που έχουν εύρος από το 0 μέχρι το 1. Στο συγκεκριμένο παράδειγμα η μέγιστη τιμή

¹Στην υποενότητα της συζήτησης των αποτελεσμάτων αναφερόμαστε και σε άλλες τεχνικές ομοιοποίησης που θα μπορούσαν να χρησιμοποιηθούν.

60Αναγνώριση Μουσικού Συνθέτη με χρήση Τεχνητών Νευρωνικών Δικτύων

βρίσκεται στο μετρητή της 5^{ης} βαθμίδας με τιμή 770. Διαιρώντας με αυτήν την τιμή οι πυκνότητες του κομματιού του παραδείγματος θα είχαν ως εξής:

Βαθμίδες μείζονος κλίμακας	Μετρητής	Πυκνότητες
1	314	0.407792207792208
#1	8	0.0103896103896104
2	481	0.624675324675325
#2	26	0.0337662337662338
3	455	0.590909090909091
4	220	0.285714285714286
#4	160	0.207792207792208
5	770	1
#5	4	0.00519480519480519
6	589	0.764935064935065
#6	46	0.0597402597402597
7	564	0.732467532467532

Πίνακας 3.2: Παράδειγμα πυκνότητας βαθμίδων μείζονος κλίμακας στο παραπάνω κομμάτι.

Στο παραπάνω παράδειγμα χρησιμοποιήσαμε τον όρο βαθμίδες της μείζονος κλίμακας με την κυριολεκτική έννοια για να γίνει πιο κατανοητό το τι θέλουμε να κάνουμε αλλά και για οικονομία χώρου. Στα πειράματα που θα κάνουμε θεωρούμε τις βαθμίδες της μείζονος κλίμακας αλλά σε ολόκληρο το εύρος των οκτάβων του κομματιού. Ένα ικανοποιητικό εύρος οκτάβων για όλα τα κομμάτια των πειραμάτων είναι οι έξι οκτάβες, με την κάθε οκτάβα να περιέχει 12 βαθμίδες. Δηλαδή οι πίνακες που χρησιμοποιούνται στα πειράματα έχουν 72 θέσεις βαθμίδων-μετρητών- πυκνοτήτων (στο παραπάνω παράδειγμα υπήρχαν οι 12 θέσεις της μίας οκτάβας).

Το ΤΝΔ που χρησιμοποιούμε

Το ΤΝΔ θα τροφοδοτηθεί με πυκνότητες των βαθμίδων της μείζονος κλίμακας στην αντίστοιχη τους τονικότητα, πράγμα που θα επιτευχθεί με τον τρόπο που περιγράψαμε παραπάνω. Οι εισοδοί δηλαδή του ΤΝΔ θα είναι 72, που όπως περιγράψαμε θα αντιπροσωπεύουν τις βαθμίδες που αντιστοιχούν στις 12 βαθμίδες (συμπεριλαμβανομένων και των αλλοιώσεων) σε 6 οκτάβες. Η n -οστή είσοδος που ικανοποιεί την σχέση $n = 1(\text{mod}(12))$ αντιπροσωπεύει την πρώτη βαθμίδα στην οκτάβα που αντιστοιχεί στη δωδεκάδα της.

Το ΤΝΔ θα εκπαιδευτεί να απαντάει στο αν οι πυκνότητες που παρουσιάστηκαν στην είσοδό του αντιστοιχούν σε ένα συγκεκριμένο συνθέτη ή όχι. Η απάντησή του θα δίνεται από τη μοναδική έξοδό του η οποία θα παίρνει τιμή 1 ή 0 αν το κομμάτι έχει συντεθεί ή όχι αντίστοιχα από το συνθέτη σπου οποιού τις πυκνότητες έχει εκπαιδευτεί.

Κατά την εκπαίδευση παρουσιάζεται στις εισόδους ένα πρότυπο πυκνοτήτων βαθμίδων ενός κομματιού και στόχος του δικτύου είναι προσαρμοστεί στην τιμή εξόδου 1 αν το κομμάτι έχει συντεθεί από το συνθέτη που το ΤΝΔ θα αναγνωρίζει, αλλιώς 0. Τα νευρωνικά δίκτυα που χρησιμοποιούμε είναι πολυεπίπεδα εμπρόσθια τροφοδότησης.

Χρησιμοποιούμε το *Neural Network Toolbox* του προγράμματος *MATLAB*. Περιγράφουμε τα δύο πειράματα τα οποία έγιναν.

3.3.4 Το πρώτο πείραμα

Το πρώτο ΤΝΔ θα εκπαιδευτεί για να αναγνωρίζει αν ένα κομμάτι έχει συντεθεί από τον Chopin ή όχι. Για τα κομμάτια που δε συντέθηκαν από τον Chopin διαλέξαμε κομμάτια από τους Beethoven και Mozart για αντιπαραδείγματα στο ΤΝΔ. Για το πρώτο πείραμα η αποστολή που έχει το δίκτυο είναι σαφής από την άποψη ότι οι Beethoven και Mozart έζησαν και έγραψαν μουσική σε διαφορετική εποχή απότι ο Chopin. Για να διατηρήσουμε όσο το δυνατόν περισσότερο τα μορφολογικά στοιχεία των κομματιών μεταξύ των τριών συνθετών ουδέτερα χρησιμοποιούμε παρτιτούρες (MIDI αρχεία) από κομμάτια ερμηνευμένα από πιάνο.

Συλλέξαμε 48 κομμάτια συντεθειμένα από τον Chopin και 34 που δεν συνετέθησαν από τον Chopin, συγκεκριμένα 13 που συνετέθησαν από τον Beethoven και 21 από τον Mozart. Τα κομμάτια για το συγκεκριμένο πείραμα τα πήραμε σε μορφή MIDI από την παρακάτω ιστοσελίδα :

<http://www.piano-midi.de/>

3.3.5 Τεχνικά στοιχεία του πρώτου πειράματος

Το άθροισμα των κομματιών που συλλέξαμε για το πρώτο πείραμα είναι 82. Χρησιμοποιούμε

- (α) 60 κομμάτια σαν σύνολο εκπαίδευσης και
- (β) 22 σαν σύνολο ελέγχου.

Τα κομμάτια που έχουν συντεθεί από τον Chopin στο σύνολο εκπαίδευσης προς τα συνολικά 60 του συνόλου εκπαίδευσης πρέπει να προσεγγίζουν το συνολικό ποσοστό των κομματιών που διαθέτουμε, δηλαδή 48 προς 82. Για να γίνουμε πιο σαφείς, αν συμβολίσουμε με x τα κομμάτια του Chopin που θέλουμε να χρησιμοποιήσουμε στο σύνολο εκπαίδευσης πρέπει να ισχύει

$$\left\lfloor \frac{x}{60} \right\rfloor = \left\lfloor \frac{48}{82} \right\rfloor$$

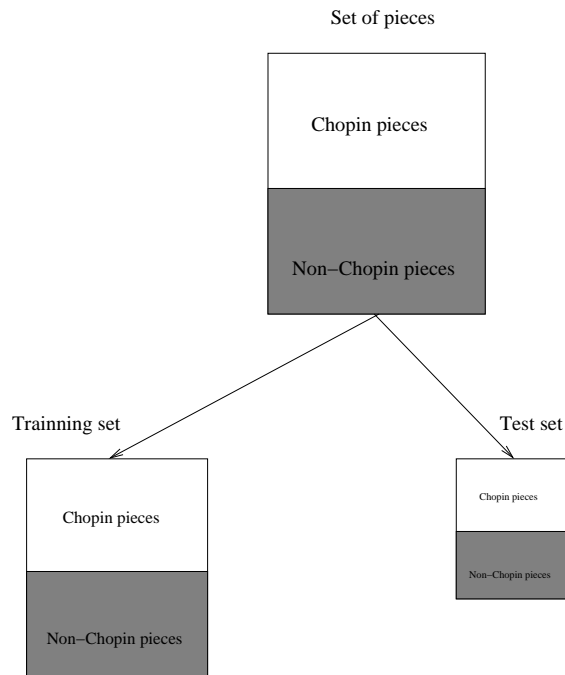
ή

$$\left\lceil \frac{x}{60} \right\rceil = \left\lceil \frac{48}{82} \right\rceil$$

κάτι που μπορούμε σχηματικά να δούμε στο πρόχειρο σχήμα 3.1.

Η διαδικασία εκπαίδευσης είναι πιο αποτελεσματική όταν οι επιθυμητές έξοδοι δεν παρουσιάζονται ομαδοποιημένες αλλά αναμεμειγμένες. Αυτό σημαίνει πως πρέπει να κατασκευάσουμε έναν πίνακα στόχων που να μην αποτελείται από συνεχόμενους άσους και μηδενικά αλλά σε τυχαία σειρά, όπως φαίνεται στο σχήμα 3.2

Επίσης πρέπει να προσέξουμε το ποσοστό κομματιών μείζονος και ελάσσονας κλίμακας της κάθε ομάδας συνθετών να μη διαφέρει σημαντικά γιατί τότε το δίκτυο θα μάθαινε να ξεχωρίζει τα κομμάτια με κριτήριο την κλίμακά τους. Για παράδειγμα



Σχήμα 3.1: Πρόχειρη σχηματική αναπαράσταση του χωρισμού των κομματιών για το πρώτο πείραμα σε σύνολο εκπαίδευσης και δοκιμής.

	Patterns	Targets	
(a)	Input(1)	1	(b)
	•	•	
	•	•	
	•	•	
	Input(i-1)	1	
	Input(i)	1	
	Input(i+1)	0	
	Input(i+2)	0	
	•	•	
	•	•	
•	•		
Input(n)	0		

	Patterns	Targets	
(a)	Input(x)	0	(b)
	•	•	
	•	•	
	•	•	
	Input(z)	0	
	Input(k)	1	
	Input(y)	0	
	Input(m)	1	
	•	•	
	•	•	
•	•		
Input(l)	1		

Σχήμα 3.2: Το σύνολο (a) έχει ομαδοποιημένους στόχους πράγμα που κάνει την εκπαίδευση του δικτύου πιο αναποτελεσματική, αντίθετα με το σύνολο (b).

αν όλα τα κομμάτια του Chopin ήταν στην ελάχισσα κλίμακα και όλα τα κομμάτια των Beethoven και Mozart στη μείζονα τότε το δίκτυο αντί για τα κομμάτια του Chopin θα αναγνώριζε τα κομμάτια σε ελάχισσα κλίμακα (ακόμα κι αν αυτά είχαν συντεθεί από τον Mozart ή τον Beethoven).

Εκπαιδεύσαμε δύο ειδών ΤΝΔ, με ένα και με δύο κρυφά στρώματα νευρώνων. Τα δίκτυα με το ένα κρυφό στρώμα νευρώνων ήταν 8 διαφορετικά δίκτυα με

$$2, 4, 6, 8, 10, 12, 14, 16$$

κρυφούς νευρώνες και τα δίκτυα με δύο κρυφά στρώματα ήταν 9 με τοπολογίες

$$(2, 2), (2, 4), (2, 6),$$

$$(4, 2), (4, 4), (4, 6),$$

$$(6, 2), (6, 4), (6, 6)$$

όπου με (x, y) συμβολίζουμε ένα ΤΝΔ με δύο κρυφά στρώματα με το πρώτο στρώμα να περιέχει x και το δεύτερο y νευρώνες.

Για καθένα από αυτά τα παραπάνω δίκτυα κάναμε 50 διαδικασίες εκπαίδευσης-δοκιμής από τις οποίες κρατήσαμε τον μέσο όρο επίδοσης καθώς και την καλύτερη επίδοση για το κάθε δίκτυο. Αναφορικά με τη διαδικασία δοκιμής, όταν στο δίκτυο παρουσιαστεί ένα πρότυπο εισόδων αυτό οφείλει να απαντήσει με 1 ή 0 αν αποφάσισε πως το κομμάτι συντέθηκε ή όχι από τον Chopin αντίστοιχα. Όμως οι απαντήσεις του δικτύου δύσκολα θα είναι οι ακέραιες τιμές 1 ή 0 λόγω της σύνθεσης των σιγμοειδών συναρτήσεων που καθορίζουν την απάντησή του. Θεωρούμε την απάντηση του δικτύου σαν 1 αν αυτή ξεπερνά ή είναι ίση με το 0.5 και 0 αλλιώς.

Όλα τα παραπάνω δίκτυα εκπαιδεύτηκαν με τη μέθοδο *Levenberg-Marquardt* [56] με ανοχή σφάλματος το 10^{-2} και ρυθμό εκπαίδευσης 10^{-3} . Η συνάρτηση ενεργοποίησης στους νευρώνες των κρυφών στρωμάτων είναι η σιγμοειδής και του νευρώνα εξόδου η ταυτοτική.

3.3.6 Αποτελέσματα του πρώτου πειράματος

Οι πίνακες 1 και 2 δείχνουν τα αποτελέσματα των δικτύων με ένα και δύο κρυφά στρώματα αντίστοιχα. Παρατηρούμε ότι η μέση τιμή της επίδοσης των 50 δικτύων κάθε είδους αυξάνεται καθώς αυξάνονται οι νευρώνες ενώ η καλύτερη επίδοση ποικίλη. Στα 22 κομμάτια του συνόλου δοκιμής το δίκτυο αποφάσισε σωστά για τα 21 πετυχαίνοντας την καλύτερη επίδοσή του, ενώ η μέση επίδοση μας λέει πως το δίκτυο μπορεί να απαντάει σωστά με ασφάλεια για περισσότερα από 17 στα 22 κομμάτια.

3.4 Το δεύτερο πείραμα

Στο πρώτο πείραμα κάποιος θα μπορούσε να πει πως υπάρχουν σημαντικές δομικές διαφορές μεταξύ των κομματιών που συνετέθησαν από τον Chopin και από τους Beethoven ή Mozart λόγω της διαφορετικότητας στον τρόπο σύνθεσης στις εποχές που έζησαν και συνέθεσαν οι παραπάνω. Αυτές οι δομικές διαφορές θα μπορούσαν να επηρεάσουν τις πυκνότητες των βαθμίδων της μείζονος κλίμακας και έτσι το πρώτο πείραμα θα μπορούσε να χαρακτηριστεί ως πείραμα αναγνώρισης εποχής σύνθεσης

64Αναγνώριση Μουσικού Συνθέτη με χρήση Τεχνητών Νευρωνικών Δικτύων

και όχι συνθέτη.

Για να ελέγξουμε αν μπορούμε να αναγνωρίσουμε το συνθέτη ενός κομματιού διεξάγουμε το δεύτερο πείραμα. Το ΤΝΔ τώρα πρέπει να αναγνωρίσει αν ένα κομμάτι έχει συντεθεί από τον Bach ή από τον Handel. Οι δύο αυτοί συνθέτες γεννήθηκαν την ίδια χρονιά, το 1685, και τα έργα τους παρουσιάζουν σημαντικές ομοιότητες, όπως έχει επισημανθεί από αρκετούς μουσικούς και μουσικολόγους. Αν ένα ΤΝΔ θα μπορούσε να αναγνωρίσει διαφορές στις πυκνότητες βαθμίδων της μείζονος κλίμακας μεταξύ κομματιών αυτών των δύο συνθετών, τότε θα μπορεί να το κάνει και για κομμάτια άλλων ζευγαριών συνθετών.

Τα κομμάτια συλλέχθηκαν από την παρακάτω ιστοσελίδα :

<http://www.classicalmidiconnection.com/>

Συγκεκριμένα, συλλέξαμε 42 κομμάτια του Bach και 41 του Handel με την αναλογία κομματιών στη μείζονα προς την ελάσσονα κλίμακα να πλησιάζει το 1 για κάθε συνθέτη. Προσπαθήσαμε τέλος τα κομμάτια να είναι ενορχηστρικά και μορφολογικά όμοια. Από τα συνολικά 83 κομμάτια

(α) τα 54 αποτελούν το σύνολο εκπαίδευσης και

(β) τα 29 το σύνολο δοκιμής.

Ο διαχωρισμός και η τυχαία ανάμειξη των κομματιών έγιναν με τον τρόπο που έγιναν και στο πρώτο πείραμα. Οι τεχνικές λεπτομέρειες που αφορούν το ΤΝΔ είναι οι ίδιες με αυτό του πρώτου πειράματος.

3.4.1 Αποτελέσματα του δεύτερου πειράματος

Η αποτελεσματικότητα του δικτύου στο δεύτερο πείραμα είναι παρόμοια με αυτή που είχαμε στο πρώτο πείραμα. Τα αποτελέσματα της μέγιστης επίδοσης είναι ενδεικτικά καθώς τα καταγεγραμμένα αποτελέσματα με άλλα που είχαμε για το ίδιο πείραμα διέφεραν. Οι μέσοι όροι τέλος μας δείχνουν ότι περισσότεροι νευρώνες σημαίνουν καλύτερα αποτελέσματα.

Πλήθος νευρώνων	Μέση επίδοση	Καλύτερη επίδοση
2	70.2727 %	86.3636 %
4	72.6364 %	81.8182 %
6	73.0000 %	86.3636 %
8	73.5455 %	86.3636 %
10	75.6364 %	95.4545 %
12	76.0000 %	86.3636 %
14	77.0000 %	90.9091 %
16	78.5455 %	95.4545 %

Πίνακας 3.3: Αποτελέσματα του πρώτου πειράματος για δίκτυα με ένα κρυφό στρώμα. Σκοπός του δικτύου είναι να αναγνωρίσει αν ένα κομμάτι έχει συντεθεί από τον Chopin ή όχι (δηλαδή από τους Beethoven ή Mozart).

3.5 Συζήτηση

Παρότι τα αποτελέσματα είναι ενδεικτικά συμπεραίνουμε πως οι πυκνότητες των βαθμίδων της μείζονος κλίμακας παίζουν κάποιο ρόλο στην αναγνώριση συνθέτη, και όχι μόνο στην αναγνώριση της εποχής σύνθεσης ενός κομματιού. Τα τεχνητά νευρωνικά δίκτυα μπορούν σε ικανοποιητικό βαθμό να εντοπίσουν κανονικότητες στις φαινομενικά ανεπαίσθητες διαφορές που παρουσιάζουν οι παραπάνω πυκνότητες στο ανθρώπινο μάτι. Οι αποφάσεις αυτές των δικτύων δεν θα μπορούσαν ακόμα να αποτελούν αυτοτελές κριτήριο, θα μπορούσαν όμως να δίνουν συμπληρωματικές ενδείξεις σε μουσικολόγους και μουσικούς που με την εμπειρία τους θα έδιναν τις κύριες κατευθύνσεις για την αναζήτηση ενός συνθέτη.

Το πρώτο πείραμα, που αφορούσε το αν ο συνθέτης ενός κομματιού ήταν ο Chopin ή όχι, αναφέραμε πως έγινε χρησιμοποιώντας παρτιτούρες κομματιών εκτελεσμένα σε πιάνο. Στα κομμάτια που έχουν γραφτεί για πιάνο συνήθως το δεξί χέρι παίζει τη μελωδία λόγω της ψηλότερης συχνότητας των νοτών που αντιστοιχούν σε αυτό το χέρι, ενώ το αριστερό συνήθως δίνει την αρμονική συνοδεία. Έγινε το ίδιο πείραμα με διαχωρισμό όμως των νοτών που χρησιμοποιούνται από το δεξί και το αριστερό χέρι, αλλά τα αποτελέσματα και στις δύο περιπτώσεις ήταν χειρότερα από αυτά που παρατίθενται σε αυτήν την εργασία, δηλαδή των νοτών και των δύο χεριών.

Καλύτερα αποτελέσματα θα μπορούσαμε να έχουμε σε αυτήν την εργασία με τη χρήση μεγαλύτερων νευρωνικών δικτύων και με καλύτερα προσαρμοσμένους αλγόριθμους εκπαίδευσης. Θα μπορούσαμε να χρησιμοποιήσουμε καλύτερες τεχνικές ομαλοποίησης και κανονικοποίησης των δεδομένων που θεωρούν επιπλέον στατιστικά στοιχεία όπως η διασπορά χρήσης βαθμίδων.

Η χρήση πυκνοτήτων *υπό σύμβαση βαθμίδων* θα μπορούσε επίσης να αποτελεί ένα βοηθητικό εργαλείο. Υπο σύμβαση βαθμίδες θα μπορούσαμε να χαρακτηρίσουμε για παράδειγμα τις βαθμίδες που ακολουθούν ή έπονται της τρίτης ή της πέμπτης βαθμίδας, ή το λόγο χρήσης της αλλοίωσης μίας βαθμίδας προς τη βαθμίδα αυτή.

Εναλλακτικά, θα μπορούσαμε να χρησιμοποιήσουμε τεχνικές ομαδοποίησης όπως ο αλγόριθμος *k-th window* με ή χωρίς τη χρήση τεχνητών νευρωνικών δικτύων.

Πλήθος νευρώνων	Μέση επίδοση	Καλύτερη επίδοση
(2,2)	67.1818 %	81.8182 %
(4,2)	69.6364 %	90.9091 %
(6,2)	68.3636 %	81.8182 %
(2,4)	66.7273 %	81.8182 %
(4,4)	67.5455 %	86.3636 %
(6,4)	68.8182 %	86.3636 %
(2,6)	62.6364 %	81.8182 %
(4,6)	68.4545 %	86.3636 %
(6,6)	72.0000 %	95.4545 %

Πίνακας 3.4: Αποτελέσματα του πρώτου πειράματος για δίκτυα με δύο κρυφά στρώματα. Σκοπός του δικτύου είναι να αναγνωρίσει αν ένα κομμάτι έχει συντεθεί από τον Chopin ή όχι (δηλαδή από τους Beethoven ή Mozart).

66Αναγνώριση Μουσικού Συνθέτη με χρήση Τεχνητών Νευρωνικών Δικτύων

Πλήθος νευρώνων	Μέση επίδοση	Καλύτερη επίδοση
2	73.8667 %	86.6667 %
4	74.8667 %	86.6667 %
6	77.0000 %	86.6667 %
8	76.2667 %	86.6667 %
10	78.0667 %	86.6667 %
12	77.6000 %	90.0000 %
14	76.9333 %	86.6667 %
16	78.8000 %	90.0000 %

Πίνακας 3.5: Αποτελέσματα του δεύτερου πειράματος για δίκτυα με ένα κρυφό στρώμα. Σκοπός του δικτύου είναι να αναγνωρίσει αν ένα κομμάτι έχει συντεθεί από τον Bach ή όχι, δηλαδή από τον Handel

Πλήθος νευρώνων	Μέση επίδοση	Καλύτερη επίδοση
(2,2)	73.0000 %	90.0000 %
(4,2)	74.9333 %	86.6667 %
(6,2)	76.5333 %	90.0000 %
(2,4)	70.0000 %	83.3333 %
(4,4)	77.3333 %	93.3333 %
(6,4)	75.7333 %	93.3333 %
(2,6)	67.0667 %	86.6667 %
(4,6)	74.2667 %	86.6667 %
(6,6)	78.0667 %	93.3333 %

Πίνακας 3.6: Αποτελέσματα του δεύτερου πειράματος για δίκτυα με δύο κρυφά στρώματα. Σκοπός του δικτύου είναι να αναγνωρίσει αν ένα κομμάτι έχει συντεθεί από τον Bach ή όχι, δηλαδή από τον Handel.

Βιβλιογραφία

- [1] I. Aleksander and H. Morton, *An introduction to neural computing*, Van Nostrand Reinhold Co., New York, NY, USA, 1990.
- [2] S. Becker, "Unsupervised learning procedures for neural networks", *International Journal of Neural Systems*, vol. 2, 17-33, 1991.
- [3] Y Bengio, P Frasconi, and P Simard, "The problem of learning long-term dependencies in recurrent networks", In the proceedings of the *Proceedings of the IEEE International Conference on Neural Networks*, IEEE Press, 1183-1195, 1993.
- [4] Y Bengio, P Simard, and P Frasconi, "Learning long-term dependencies with gradient descent is difficult", *IEEE Transactions on Neural Networks*, , No. 5, 157-166, 1994.
- [5] Jan Beran, *Statistics in Musicology (Interdisciplinary Statistics.)*, Chapman & Hall/CRC, 1st ed., Ουλψ 2003.
- [6] Jan Beran, "Music - chaos, fractals and information", *Chance Vol. 17*, , No. 4, 2004.
- [7] J. Beran, "Music, chaos, fractals and information", *Chance, Vol. 17*, , No. 4, 7-16, 2004.
- [8] J J Bharucha, "Pitch, harmony, and neural nets: A psychological perspective", In the proceedings of the *In*, MIT Press/Bradford Books, 84-99, 1991.
- [9] J J Bharucha and P M Todd, "Modeling the perception of tonal structure with neural nets", *Computer Music Journal*, , No. 13, 44-53, 1989.
- [10] J S Bridle, "Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters", In the proceedings of the *Advances in Neural Information Processing Systems 2*, Morgan Kaufmann, 211-217, 1990.
- [11] D Burr and Y Miyata, "Hierarchical recurrent networks for learning musical structure", In the proceedings of the *In*, 1993.
- [12] G W Cottrell, "Learning simple arithmetic procedures", *Connection Science*, , No. 5, 37-58, 1993.
- [13] J Denker, D Schwartz, B Wittner, S Solla, R Howard, L Jackel, and J Hopfield, "Large automatic learning, rule extraction, and generalization", *Complex Systems*, , No. 1, 877-922, 1987.
- [14] C Dodge and T A Jerse, "Computer music: Synthesis, composition, and performance", 1985.
- [15] D. Eck and J. Schmidhuber, "A first look at music composition using lstm recurrent neural networks", *Technical Report*, 2002.

- [16] J L Elman, "Finding structure in time", *Cognitive Science*, , No. 14, 179-216, 1990.
- [17] J L Elman, "Learning and development in neural networks: The importance of starting small", *Cognition*, , No. 48, 71-99, 1993.
- [18] P Fraisse, "Rhythm and tempo", In the proceedings of the *The Psychology of Music*, Academic Press, 149-180, 1982.
- [19] J. A. Franklin, "Recurrent neural networks for music computation", .
- [20] Simon Haykin, *Adaptive filter theory (3rd ed.)*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.
- [21] Simon Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1998.
- [22] G E Hinton, "Learning distributed representations of concepts", In the proceedings of the *In Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, Erlbaum, 1-12, 1986.
- [23] A. J. Hsu and A. J. Hsu, "Fractal geometry of music", *Proc. Natl. Acad. Sci. USA*, Vol. 87, 938-941, 1990.
- [24] K Jones, "Compositional applications of stochastic processes", *Computer Music Journal*, , No. 5, 45-61, 1981.
- [25] M R Jones and M Boltz, "Dynamic attending and responses to time", *Psychological Review*, , No. 96, 459-491, 1989.
- [26] T Kohonen, "A self-learning musical grammar.associative memory of the second kind", In the proceedings of the *In Proceedings of the 1989 International Joint Conference on Neural Networks*, 1-5, 1989.
- [27] T Kohonen, P Laine, K Tiits, and K Torkkola, "A nonheuristic automatic composing method", In the proceedings of the *In*, MIT Press/Bradford Books, 229-242, 1991.
- [28] C L Krumhansl, J J Bharucha, and E J Kessler, "Perceived harmonic structure of chords in three related musical keys", *Journal of Experimental Psychology: Human Perception and Performance*, , No. 8, 24-36, 1982.
- [29] C L Krumhansl and E J Kessler, "Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys", *Psychological Review*, , No. 89, 334-368, 1982.
- [30] B Laden and D H Keefe, "The representation of pitch in a neural net model of chord classification", *Computer Music Journal*, , No. 13, 12-26, 1989.
- [31] J P Lewis, "Creation by refinement and the problem of algorithmic music composition", In the proceedings of the *In*, MIT Press/Bradford Books, 212-228, 1991.
- [32] H C Longuet-Higgins, "The perception of melodies", In the proceedings of the *Nature*, British Psychological Society/MIT Press, 646-653, 1976.
- [33] H C Longuet-Higgins, "The perception of music (review lecture)", *Proceedings of the Royal Society of London*, , No. 205, 307-332, 1979.

- [34] D Lorrain, "A panoply of stochastic 'cannons'", *Computer Music Journal*, , No. 3, 48–55, 1980.
- [35] D G Loy, "Connectionism and musiconomy", In the proceedings of the *Music and Connectionism*, MIT Press, 23–36, 1991.
- [36] Shepherd G. M. and Koch C., "Introduction to synaptic circuits", In *The Synaptic Organization of the Brain*, edited by G. M. Shepherd, Oxford University Press, New York, NY, 3–31, 1990.
- [37] G.D. Magoulas, V.P. Plagianakos, G.S. Androulakis, and M.N. Vrahatis, "A framework for the development of globally convergent adaptive learning rate algorithms", *International Journal of Computer Research*, vol. 1, 1–10, 2001.
- [38] G.D. Magoulas, M.N. Vrahatis, and G.S. Androulakis, "Effective backpropagation training with variable stepsize", *Neural Networks*, vol. 10, No. 1, 69–82, 1997.
- [39] George D. Magoulas, Michael N. Vrahatis, and George S. Androulakis, "Effective backpropagation training with variable stepsize", *Neural Netw.*, vol. 10, No. 1, 69–82, 1997.
- [40] George D. Magoulas, Michael N. Vrahatis, and G. S. Androulakis, "Improving the convergence of the backpropagation algorithm using learning rate adaptation methods", *Neural Computation*, vol. 11, No. 7, 1769–1796, 1999.
- [41] J L McClelland, "On the time relations of mental processes: An examination of systems of processes in cascade", *Psychological Review*, , No. 86, 287–330, 1979.
- [42] J. M. Mendel and R. W. McLaren, "Reinforcement-learning control and pattern recognition systems", , 287–318, 1994.
- [43] Marvin Minsky, "Steps toward artificial intelligence", , 406–450, 1995.
- [44] M C Mozer, "Rambot: A connectionist expert system that learns by example", In the proceedings of the *In M. Caudill and C. Butler (Eds.), Proceedings of the IEEE First Annual International Conference on Neural Networks*, IEEE Publishing Services, 693–700, 1987.
- [45] M C Mozer, "A focused back-propagation algorithm for temporal pattern recognition", *Complex Systems*, , No. 3, 349–381, 1989.
- [46] M C Mozer, "Induction of multiscale temporal structure", In the proceedings of the *Advances in Neural Information Processing Systems 4*, Morgan Kaufmann, 275–282, 1992.
- [47] M. C. Mozer, "Neural network music composition by prediction: Exploring the benefits of psycho acoustic constraints and multiscale processing", *Connection Science*, 1994.
- [48] M C Mozer and S Das, "A connectionist symbol manipulator that discovers the structure of context-free languages", In the proceedings of the *In*, Morgan Kaufmann, 863–870, 1993.
- [49] C Myers, "Learning with delayed reinforcement through attention-driven buffering (technical report)", *Journal of Neural Systems*, , No. 1, 337–346, 1990.

- [50] B. K. Natarajan, editor, *Machine Learning: A Theoretical Approach*, Morgan Kaufmann, San Mateo, CA, 1991.
- [51] D. B. Parker, “Optimal algorithms for adaptive networks: Second order back-propagation, second order direct propagation, and second order hebbian learning”, In the proceedings of the *In Proceedings of the IEEE International Conference on Neural Networks*, vol. 2, 593-600, 1987.
- [52] B A Pearlmutter, “Learning state space trajectories in recurrent neural networks”, *Neural Computation*, , No. 1, 263-269, 1989.
- [53] F J Pineda, “Generalization of back-propagation to recurrent neural networks”, *Physical Review Letters*, , No. 59, 2229-2232, 1987.
- [54] V.P. Plagianakos, M.N. Vrahatis, and G.D. Magoulas, “Nonmonotone methods for backpropagation training with adaptive learning rate”, In the proceedings of the *IEEE International Joint Conference on Neural Networks (IJCNN'99)*, CD-ROM Proceedings, Paper No.2001, Session 5.1, Washington, U.S.A., 1999.
- [55] M B Ring, “Learning sequential tasks by incrementally adding higher orders”, In the proceedings of the *Advances in Neural Information Processing Systems 5*, Morgan Kaufmann, 115-122, 1993.
- [56] S. Roweis, “Levenberg marquardt optimization”, In the proceedings of the *Numerical Recipes in C, Second Edition*, , παγες = 683-685, πύβλισηερ = .
- [57] D E Rumelhart, G E Hinton, and R J Williams, “Learning internal representations by error propagation”, In the proceedings of the *In: Rumelhart, D.E., McClelland, J.L. (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, 1986.
- [58] J Schmidhuber, “Learning unambiguous reduced sequence descriptions”, In the proceedings of the *Advances in Neural Information Processing Systems 4*, Morgan Kaufmann, 291-298, 1992.
- [59] J H Schmidhuber, M C Mozer, and D Prelinger, “Continuous history compression”, In the proceedings of the *Proc. of Intl. Workshop on Neural Networks, RWTH Aachen*, Augustinus, 87-95, 1993.
- [60] R N Shepard, “Geometrical approximations to the structure of musical pitch”, *Psychological Review*, , No. 89, 305-333, 1982.
- [61] R N Shepard, “Toward a universal law of generalization for psychological science”, *Science*, , No. 237, 1317-1323, 1987.
- [62] C Stevens and J Wiles, “Representations of tonal music: A case study in the development of temporal relationship”, In the proceedings of the *Proceedings of the 1993 Connectionist Models Summer School*, Erlbaum, 228-235, 1994.
- [63] P M Todd, “A connectionist approach to algorithmic composition”, *Computer Music Journal*, , No. 13, 27-43, 1989.
- [64] P M Todd, “A connectionist approach to algorithmic composition”, In the proceedings of the *Music and Connectionism*, Cambridge, MA, MIT Press Bradford Books, 173-194, 1991.

-
- [65] M. N. Vrahatis, G. S. Androulakis, J. N. Lambrinos, and G. D. Magoulas, “A class of gradient unconstrained minimization algorithms with adaptive stepsize”, *J. Comput. Appl. Math.*, vol. 114, No. 2, 367–386, 2000.
- [66] Michael N. Vrahatis, George D. Magoulas, and Vassilis P. Plagianakos, “Globally convergent modification of the quickprop method”, *Neural Process. Lett.*, vol. 12, No. 2, 159–170, 2000.
- [67] Bernard Widrow and Marcian E. Hoff, “Adaptive switching circuits”, , 123–134, 1988.
- [68] Bernard Widrow and Samuel D. Stearns, *Adaptive signal processing*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1985.