



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΥΠΟΛΟΓΙΣΤΩΝ

Αυτόματη Αλλαγή Μουσικού Είδους σε
Συμβολική Μορφή Αναπαράστασης με Χρήση
Βαθιών Νευρωνικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΟΡΦΕΩΣ ΚΟΦΙΝΑΚΟΥ

Επιβλέπων: Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2021



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής & Υπολογιστών

Αυτόματη Αλλαγή Μουσικού Είδους σε Συμβολική Μορφή Αναπαράστασης με Χρήση Βαθιών Νευρωνικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΟΡΦΕΩΣ ΚΟΦΙΝΑΚΟΥ

Επιβλέπων: Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 15η Ιουλίου 2021.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.

.....
Α-Γ Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

.....
Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2021



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής & Υπολογιστών

Copyright ©–All rights reserved Ορφεύς Κοφινάκος, 2021.

Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

(Υπογραφή)

.....
ΟΡΦΕΥΣ ΚΟΦΙΝΑΚΟΣ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

© 2021– All rights reserved

Περίληψη

Η εκτεταμένη χρήση της Τεχνητής Νοημοσύνης στη μουσική, όπως και σε όλα τα πεδία καλλιτεχνικής δημιουργίας, έχει οδηγήσει σε αυξημένη ζήτηση από τους δημιουργούς για εργαλεία που θα επιτρέψουν την ακόμη πιο άμεση και ολοκληρωμένη παραγωγή έργων. Ένα δυνητικά πραγματοποιήσιμο τέτοιο εργαλείο είναι η μεταφορά μουσικών κομματιών από είδος σε είδος. Η διερεύνηση των πολλαπλών συνεργειών της Τεχνητής Νοημοσύνης με τη μουσική και ειδικότερα αυτού του εγχειρήματος έχει γίνει συχνό αντικείμενο έρευνας στο παρελθόν, με τη χρήση ηχητικών αρχείων. Η προσέγγιση που παρουσιάζεται σε αυτήν την εργασία διαφοροποιείται από τις προηγούμενες στο βαθμό που επιλέγει να εργαστεί με αρχεία συμβολικών δεδομένων (MIDI). Η επιλογή αυτή επιβάλλει την ανάλυση της μουσικής σε επίπεδο σύνθεσης, καθώς μεταβλητές που εμφανίζονται στα ηχητικά αρχεία, όπως το ηχόχρωμα, απουσιάζουν. Η διαδικασία που ακολουθήσαμε ήταν να επιλέξουμε απλοποίηση της κωδικοποίησης πληροφοριών σε μορφή επαρκή για την εκπαίδευση νευρωνικού δικτύου με αρχιτεκτονικές βαθιάς μηχανικής μάθησης. Οργανώσαμε πειράματα με ένα γενικό μοντέλο autoencoder με χρήση ξεχωριστών decoder για κάθε είδος και έναν καθολικό encoder. Το μοντέλο ακολουθεί το πρωτόκολλο ενός συνηθισμένου encoder-decoder model το οποίο αποτελεί τρόπο οργάνωσης επαναλαμβανόμενων νευρωνικών δικτύων (RNN) για χρήση σε προβλήματα πρόβλεψης sequence-to-sequence με χρήση LSTM με πολλαπλά layers, για εκπαίδευση σε μουσικά κομμάτια από το Lakh Pianoroll Dataset. Χρησιμοποιήθηκε στρατηγική Teacher Forcing, προκειμένου να βελτιστοποιηθεί η παραγωγή χρήσιμων αποτελεσμάτων. Το μοντέλο εφαρμόστηκε με κοινά γενικά χαρακτηριστικά αρχιτεκτονικής σε δύο μορφές αναπαράστασης. Παρήχθησαν πειράματα, τα οποία αξιολογήθηκαν με ταξινομητή, ενώ όσα έδωσαν ενδιαφέροντα αποτελέσματα υπεβλήθησαν και σε ανθρώπινη αξιολόγηση. Κατά την ολοκλήρωση της εργασίας προέκυψαν σφάλματα, τα οποία εξηγούνται από την ποιότητα και την οργάνωση του dataset αλλά και τού εύρους των επιλεγμένων μουσικών ειδών προς ανάλυση. Προέκυψε το συμπέρασμα ότι το εγχείρημα έχει εμφανή περιθώρια βελτίωσης, ενώ παρέχει χρήσιμες πληροφορίες για στοχευμένες μελλοντικές έρευνες που αργά ή γρήγορα θα οδηγήσουν στην παραγωγή τού εύχρηστου εργαλείου μεταποίησης το οποίο θελήσαμε να μελετήσουμε.

Λέξεις Κλειδιά

Τεχνητά Νευρωνικά Δίκτυα, Μεταφορά Είδους, Μεταφορά Ύφους, Ανάκτηση Μουσικής Πληροφορίας, Συμβολική Μουσική Αναπαράσταση, Αναγνώριση Τονικής, Αναδρομικά Νευρωνικά Δίκτυα, Ρυθμός, Νότες

Abstract

The extensive use of AI in music, as in all fields of artistic creation, has led to increased demand from creators for tools that will allow for even more direct and complete production of works. A potentially feasible tool is to transfer music tracks from genre to genre. The research on the multiple synergies of AI with music and in particular this project (genre to genre) has become a frequent subject of research in the past, using audio files. The approach presented in this thesis differs from the previous ones to the extent that it chooses to work with symbolic data files (MIDI). This option requires the analysis of music at the synthesis level, as variables that appear in sound files, such as timbre, are absent. The process we followed was to choose to simplify the encoding of information in a format sufficient to train a neural network with deep machine learning architectures. We organized experiments with a generic autoencoder model using separate decoders for each genre and a universal encoder. The model follows the protocol of an ordinary encoder-decoder model which is a way of organizing repetitive neural networks (RNN) for use in sequence-to-sequence prediction problems using LSTM with multiple layers, by training with music tracks from Lakh Pianoroll Dataset. A Teacher Forcing strategy was used to optimise the production of useful results. The model was applied with common general architecture characteristics in two forms of representation. Experiments were produced, which were evaluated through a classifier, while those that gave interesting results were also subjected to human evaluation. Errors occurred at the completion of the work, which are explained by the quality and organization of the dataset and the wide range of the selected musical genres analyzed. It has been concluded that the project has obvious room for improvement, while providing useful information for targeted future research that will sooner or later lead to the production of the easy-to-use processing tool that we endeavoured to study.

Keywords

Artificial Neural Networks, Genre Transfer, Style Transfer, Music Information Retrieval, Symbolic Music Representation, Root Note Recognition, Recurrent Neural Networks, Rythm, Notes

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον Διπλωματικό Γιώργο Φιλανδριανό και τον επιβλέποντα καθηγητή κ. Στάμου για την ευκαιρία που μου δόθηκε να διευρύνω τις γνώσεις μου μέσω της εκπόνησης αυτής της διπλωματικής και την καθοδήγησή τους κατά τη διάρκεια ολοκλήρωσής της.

Με αφορμή την ολοκλήρωση των σπουδών μου, θα ήθελα επίσης να ευχαριστήσω τους γονείς μου για την σιωπηλή υπομονή και στήριξη τους καθώς και όλη μου την οικογένεια και τους κοντινούς μου ανθρώπους για τη βοήθειά τους αυτά τα χρόνια. Θα ήθελα επίσης να ευχαριστήσω τους συμφοιτητές και φίλους μου Παναγιώτη Προμπονά και Γιάννη Ζαρκάδα χωρίς τη βοήθεια των οποίων δεν θα μπορούσα να ολοκληρώσω αυτή τη σχολή, καθώς και τους Γιάννη Σεισόπουλο, Κυριάκο Λωτίδη και Γρηγόρη Βελέγκα για τις αξέχαστες συνεργασίες μας.

Περιεχόμενα

Περίληψη	1
Abstract	3
Ευχαριστίες	5
Περιεχόμενα	8
Κατάλογος Σχημάτων	9
1 Εισαγωγή	11
1.1 Πρόλογος - Εγχείρημα - Σχετική Έρευνα	11
1.2 Διάρθρωση διπλωματικής εργασίας	12
I Θεωρητικό Μέρος	13
2 Μουσική	15
2.1 Είδη και στυλ μουσικής	15
3 Midi	17
3.1 Εισαγωγή	17
3.2 Τρόπος λειτουργίας	19
3.3 Pianoroll	20
4 Εγχείρημα	21
4.1 Προσέγγιση με βάση τη δομή του MIDI και τον ορισμό του στυλ	21

II	Πρακτικό Μέρος	23
5	Οργάνωση δεδομένων	25
5.1	Lakh labels	25
5.2	Κωδικοποίηση πληροφοριών	26
6	Αρχιτεκτονική Μοντέλου	31
6.1	Autoencoder με πολλαπλούς decoder και κοινό encoder	31
6.2	Seq2seq	32
6.3	Teacher Forcing	33
7	Προετοιμασία και Υπερπαράμετροι	35
7.1	Συναρτήσεις ενεργοποίησης	35
7.2	Συναρτήσεις Κόστους	35
7.3	Βελτιστοποιητές	36
7.4	Μετρικές	36
8	Πειράματα	37
8.1	Συνεχόμενες τιμές Offset, Quarterlength - One Hot Encoded Pitch	37
8.2	One Hot Encoding των τριών μεταβλητών	38
8.2.1	Πρώτο στάδιο - Εκπαίδευση με ανεπεξέργαστα δεδομένα	38
8.2.2	Δεύτερο στάδιο - Εκπαίδευση μετά από μεταφορά τονικότητας	40
8.2.3	Τρίτο στάδιο - Εκπαίδευση μετά από απαλοιφή του Pitch	42
8.3	Επιπλέον αξιολόγηση	43
9	Υλοποίηση	45
10	Αποτελέσματα - Συμπεράσματα	47
10.1	Σφάλματα	47
10.2	Συμπεράσματα	48
10.3	Μελλοντική Έρευνα	48
	Βιβλιογραφία	51

Κατάλογος Σχημάτων

3.1	Παράδειγμα τυπικού MIDI συστήματος	17
3.2	Παράδειγμα μηνύματος MIDI	19
3.3	Παράδειγμα pianoroll	20
5.1	Αναπαράσταση τιμών quarterlength σε OHE πίνακες μετά από αντιστοίχιση σε ακεραίους	28
6.1	Γραφική απεικόνιση του μοντέλου	31
6.2	Γραφική απεικόνιση των ακολουθιών	32
6.3	Στάδιο decoding με και χωρίς χρήση Teacher Forcing	34
8.1	Μετρικές για το μοντέλο Rock-Jazz του πρώτου σταδίου πειραμάτων	39
8.2	Ποσοστά ταξινόμησης πρώτου σταδίου εκπαίδευσης σε ανεπεξέργαστα δεδομένα	40
8.3	Ποσοστά ταξινόμησης δεύτερου σταδίου εκπαίδευσης σε δεδομένα κοινού τσινικού κέντρου	41
8.4	Ποσοστά ταξινόμησης τρίτου σταδίου εκπαίδευσης σε δεδομένα με απαλοιφή του pitch	43

1.1 Πρόλογος - Εγχείρημα - Σχετική Έρευνα

Οι μέθοδοι Τεχνητής Νοημοσύνης (AI) προσφέρουν νέες δυνατότητες στην τέχνη και τη ψυχαγωγία, επιτρέποντας ουσιαστικές και διαδραστικές εμπειρίες. Συγχρόνως, καθώς ο κλάδος εισάγει νέα πεδία καλλιτεχνικής έκφρασης, η ίδια η βασισμένη σε AI τέχνη συνιστά μια αυτοτελή περιοχή έρευνας, θέτοντας και απαντώντας σε νέα ερευνητικά ερωτήματα τα οποία δεν θα είχαν τεθεί χωρίς την εφαρμογή AI πρακτικών στα πλαίσια της τέχνης και της ψυχαγωγίας[1]. Ο κλάδος Τεχνητή Νοημοσύνη και Μουσική (AIM) έχει αποτελέσει σύνηθες θέμα σε συνέδρια και σεμινάρια[2], και η επίδρασή του είναι ήδη αισθητή στη μουσική βιομηχανία παγκοσμίως, καθώς καλύπτει πολλές σημαντικές περιοχές του πεδίου.

Αν και έχουν περάσει πάνω από 60 χρόνια από τα πρώτα πειράματα με αλγοριθμική σύνθεση μουσικής AI, συνολικά θεωρείται από πολλούς ειδικούς στη μουσική και ερευνητές ότι ακόμα βρισκόμαστε στις πολύ πρώιμες μέρες αυτής της τεχνολογίας[3]. Τρέχουσες έρευνες χρησιμοποιούν την εφαρμογή Τεχνητής Νοημοσύνης στη μουσική σύνθεση, απόδοση, θεωρία και ψηφιακή επεξεργασία ήχου.

Μέσα σε αυτήν την πραγματικότητα, παρουσιάζει ιδιαίτερο χρηστικό ενδιαφέρον η δημιουργία ενός εργαλείου μεταποίησης μεταξύ εύχρηστων αρχείων συμβολικής αναπαράστασης μουσικής, έτσι ώστε να προσφέρεται η δυνατότητα συνθετικής ποικιλίας στην διαδικασία μουσικής παραγωγής, η οποία στρέφεται όλο και περισσότερο στην εκμετάλλευση τεχνολογικών μέσων. Η παρούσα εργασία στοχεύει στην διερεύνηση αυτής της δυνατότητας.

Η προσέγγιση της παρούσας εργασίας στηρίζεται στο συνδυασμό ήδη εφαρμοσμένων πρακτικών. Αν και η αλλαγή στυλ και είδους μουσικής για αρχεία audio έχει αποτελέσει αντικείμενο έρευνας στο παρελθόν, η αναπαράσταση των δεδομένων σε μορφή MIDI στην οποία στηρίζεται η παρούσα εργασία, διαφοροποιεί σημαντικά το αναμενόμενο αποτέλεσμα. Η ανάλυση της μουσικής γίνεται σε επίπεδο σύνθεσης, σε παρόμοια λογική με αλγόριθμους μουσικής σύνθεσης που κάνουν χρήση του πρωτοκόλλου MIDI ως συμβολικής αναπαράστασης. Η προσέγγιση αυτή αγνοεί το ηχόχρωμα και άλλα μουσικά χαρακτηριστικά που απουσιάζουν σε μια τέτοια αναπαράσταση και στα οποία βασίζονται αντίστοιχα εγχειρήματα για την μεταφορά μεταξύ ειδών μουσικής.

1.2 Διάρθρωση διπλωματικής εργασίας

Στο δεύτερο και τρίτο κεφάλαιο γίνεται μια περιεκτική ανάλυση θεωρητικών και τεχνικών ορολογιών τις οποίες χρησιμοποιεί η παρούσα εργασία. Συγκεκριμένα, δίνονται ορισμοί για το μουσικό είδος και το στυλ, όροι των οποίων η ερμηνεία ποικίλει, ως επαναλαμβανόμενες και κομβικές έννοιες για την εργασία. Ακολουθεί μια συνοπτική παρουσίαση του πρωτοκόλλου MIDI, γύρω από το οποίο έχει δομηθεί τεχνικά το εγχείρημα. Με βάση αυτές τις πληροφορίες, στο τέταρτο κεφάλαιο γίνεται μια συγκεκριμενοποίηση της προσέγγισης που ακολουθείται σε πρακτικό επίπεδο. Το πέμπτο κεφάλαιο περιγράφει τη διαδικασία συλλογής δεδομένων και την επεξεργασία που δέχθηκαν κατά την προετοιμασία τους για χρήση. Στο έκτο και έβδομο κεφάλαιο παρουσιάζεται η αρχιτεκτονική του μοντέλου νευρωνικών δικτύων που χρησιμοποιήθηκε, καθώς και οι παράμετροί του και οι λόγοι για τους οποίους προτιμήθηκαν τους. Εντός του ογδόου κεφαλαίου, αναλύονται οι διαφορετικές εκφάνσεις των πειραμάτων και τα αποτελέσματα που επέφεραν. Μετά από μια ανάλυση των τεχνικών παραμέτρων της υλοποίησης που πραγματοποιήθηκε στο ένατο κεφάλαιο, παρατίθενται στο δέκατο κεφάλαιο πιθανά σφάλματα της προσέγγισης και συμπεράσματα πάνω στα αποτελέσματα που προέκυψαν.

Μέρος Ι
Θεωρητικό Μέρος

2.1 Είδη και στυλ μουσικής

Το είδος είναι ένας όρος που αντικατοπτρίζει την τάση να ταξινομηθούν και να κατηγοριοποιηθούν έργα τέχνης που χρονολογείται από την εποχή του Έλληνα φιλόσοφου Αριστοτέλη. Ως αποτέλεσμα αυτού, στους τίτλους έργων εμφανίζονται συχνά γενικοί όροι (για παράδειγμα, συμφωνία ή κοντσέρτο). Το είδος αναγκαστικά θέτει ένα σύνολο κωδικών και προσδοκιών και ως εκ τούτου μπορεί να εκληφθεί ως κάτι που επιβάλλεται στη μουσική από μουσικές κουλτούρες, επηρεάζοντας τον τρόπο με τον οποίο γράφεται. Για παράδειγμα, οι συνθέτες και οι δημοφιλείς μουσικοί συχνά πειραματίζονται με τα συμβατικά όρια ειδών αναζητώντας νέα ακούσματα. Η πληροφορία επί του είδους μπορεί επίσης να υποδεικνύει τρόπους ακρόασης. Σύμφωνα με τον εθνομουσικολόγο William Hanks, όταν θεωρείται ως αλληλεπίδραση μεταξύ κοινωνικών δομών και μουσικού περιεχομένου, το είδος μπορεί να θεωρηθεί ότι προσφέρει (1) ένα πλαίσιο που μπορεί να χρησιμοποιήσει ένας ακροατής για να προσανατολιστεί· (2) διαδικασίες για την ερμηνεία της μουσικής· και (3) ένα σύνολο προσδοκιών (Hanks 1987).

Η έννοια του στυλ αναφέρεται σε έναν τρόπο έκφρασης, την μέθοδο με την οποία διατυπώνονται οι μουσικές κινήσεις. Υπό αυτό το πρίσμα, μπορεί να φανεί ότι σχετίζεται με την έννοια της ταυτότητας. Στη μουσική, ο ορισμός στυλ απαιτεί την εξέταση τεχνικών χαρακτηριστικών (μελωδία, χροιά, ρυθμός και αρμονία) και αφορά τρόπους με τους οποίους αυτά τα χαρακτηριστικά λειτουργούν ανεξάρτητα ή σε συνδυασμό, ή ως κατηγορίες, όπως η αντίστιξη. Με την ευρύτερη έννοια του, το στυλ μπορεί να αναφέρεται στη μουσική ως στυλ τέχνης, ενώ με τη στενότερη έννοια μπορεί να εφαρμοστεί σε μια μόνο νότα, η οποία μπορεί να έχει στυλιστικά χαρακτηριστικά που καθορίζονται από τον τόνο, τη δυναμική, τη χροιά και ούτω καθεξής. Το στυλ μπορεί να προσδιορίσει ιστορικές περιόδους, και έχει μια αντανακλαστική σχέση με τη μορφή, τη λειτουργία και το είδος μιας εργασίας. Ο όρος προέκυψε από τις τάσεις κατηγοριοποίησης του Διαφωτισμού, οι οποίες παρουσιάζονται στο έργο των Γερμανών θεωρητικών μουσικής, όπως οι Athanasius Kircher και Johann Mattheson. Ο Kircher, στο *Musurgiauniversalis* του 1650, και ο Mattheson, στο *Der Vollkommene Capellmeister* του 1739 μίλησαν για τα εθνικά στυλ. Καθώς προέκυπτε μεγαλύτερο ενδιαφέρον για τη μουσική θεωρία, έτσι με τη σειρά της προέκυψε η επιθυμία να προσδιοριστούν θεωρίες βασισμένες σε

κοινές παραδοχές.

Μια σύγχυση που μπορεί μερικές φορές να προκύψει, είναι η αλληλεπικάλυψη μεταξύ των όρων *στυλ* και *είδος* που συναντώνται στη μουσικολογική γραφή. Ο Allen Moore σημείωσε ότι, σε γενικές γραμμές, οι δημοφιλείς μουσικές μελέτες χρησιμοποιούν τον όρο *είδος*, ενώ στη γραφή για την κλασική μουσική έτεινε, τουλάχιστον μέχρι τα μέσα της δεκαετίας του 1980, να εναλλάσσεται ως όρος ελεύθερα το *είδος* και το *στυλ* (Moore 2001b). Από τα μέσα της δεκαετίας του 1980, ωστόσο, οι μουσικολόγοι τείνουν να χρησιμοποιούν τον όρο *είδος* για να περιγράψουν τις εξωτερικές, κοινωνικά διαμορφωμένες πτυχές ενός έργου και το *στυλ* προορίζεται για την εξέταση των επίσημων, εσωτερικών χαρακτηριστικών[4].

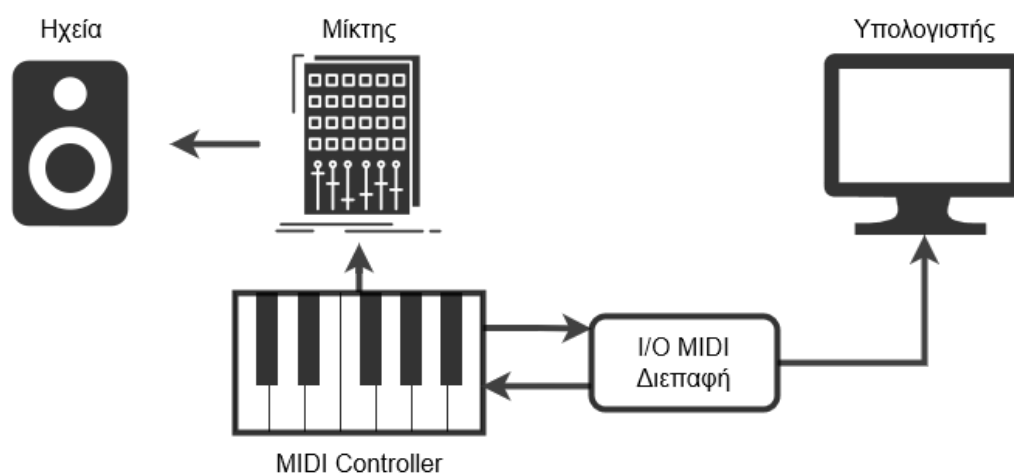
Ο Franco Fabbri ορίζει το μουσικό είδος ως «μια μορφή μουσικής, όπως αναγνωρίζεται από μια κοινότητα για οποιονδήποτε λόγο ή σκοπό ή κριτήριο, δηλαδή, ένα σύνολο μουσικών εκδηλώσεων των οποίων η πορεία διέπεται από κανόνες (οποιοδήποτε είδους) αποδεκτούς από μια κοινότητα» (Fabbri 1999). Ο Fabbri στη συνέχεια ορίζει το μουσικό *στυλ* ως «μια επαναλαμβανόμενη διάταξη χαρακτηριστικών σε μουσικές εκδηλώσεις που είναι χαρακτηριστική ενός ατόμου (συνθέτης, ερμηνευτής), μιας ομάδας μουσικών, ενός είδους, ενός τόπου, μιας χρονικής περιόδου» (Fabbri 1999). Το μουσικό είδος μπορεί επομένως να θεωρηθεί ότι είναι κάπως ευρύτερο και πιο υποκειμενικό από το *στυλ* από μια οπτική που βασίζεται στο περιεχόμενο, γεγονός που καθιστά την ταξινόμηση του είδους πιο δύσκολη αλλά και πιο ενδιαφέρουσα από την ταξινόμηση του *στυλ*. Ως αποτέλεσμα προκύπτει το φαινόμενο να αποδίδονται περισσότερα του ενός είδη σε κάποια κομμάτια μουσικής.

Θα πρέπει να σημειωθεί ότι οι διακρίσεις που έγιναν παραπάνω είναι για λόγους αποσαφήνισης της χρήσης τους σε αυτή τη διατριβή και ότι δεν εκφράζουν κάποια καθολική συμφωνία στη μουσική κοινότητα[5].

3.1 Εισαγωγή

Το Musical Instrument Digital Interface (MIDI) είναι μια γλώσσα ψηφιακών επικοινωνιών και προδιαγραφές συμβατότητας που επιτρέπουν σε πολλαπλά ηλεκτρονικά όργανα υλικού και λογισμικού, ελεγκτές επιδόσεων, υπολογιστές και άλλες σχετικές συσκευές να επικοινωνούν μεταξύ τους μέσω ενός συνδεδεμένου δικτύου (Εικόνα 1.1). Το MIDI χρησιμοποιείται για τη μετάφραση συμβάντων σχετικά με την απόδοση ή τον έλεγχο (όπως ο χειρισμός αρμονίου, η ενεργοποίηση ενός προσχεδιασμένου οπτικού εφέ, κλπ.) σε ισοδύναμα ψηφιακά μηνύματα και στη συνέχεια για τη μετάδοση αυτών των μηνυμάτων σε άλλες συσκευές MIDI όπου μπορούν να χρησιμοποιηθούν για τον έλεγχο γεννητριών ήχου και άλλων παραμέτρων επίδοσης. Μία από τις βασικές χρησιμότητες του MIDI είναι ότι τα δεδομένα του μπορούν να καταγραφούν σε μια συσκευή υλικού ή λογισμικό πρόγραμμα (γνωστό ως sequencer), όπου μπορούν να υποβληθούν σε επεξεργασία και να μεταδοθούν σε ηλεκτρονικά όργανα ή άλλες συσκευές για τη δημιουργία μουσικής ή τον έλεγχο διαφόρων παραμέτρων.

Το MIDI δεν μεταδίδει ήχο ούτε μπορεί να δημιουργήσει ήχους. Πρόκειται για μια ψηφιακή γλώσσα που δίνει εντολή σε μια συσκευή ή πρόγραμμα να δημιουργήσει, να αναπαράξει ή να



Σχήμα 3.1: Παράδειγμα τυπικού MIDI συστήματος

αλλάξει ήχους. Το MIDI είναι ένα πρωτόκολλο δεδομένων που μεταδίδει σήμα ενεργοποίησης / απενεργοποίησης και διαθέτει ένα ευρύ φάσμα παραμέτρων για να δώσει εντολή σε ένα όργανο ή συσκευή να δημιουργήσει, να αναπαράξει ή να ελέγξει ήχους ή λειτουργίες συναφείς με την παραγωγή. Λόγω αυτής της ειδοποιού διαφοράς, η διαδρομή δεδομένων MIDI και οι διαδρομές δρομολόγησης ήχου είναι εντελώς ξεχωριστές μεταξύ τους (Σχήμα 1.2). Ακόμα κι αν μοιράζονται ψηφιακά το ίδιο καλώδιο μετάδοσης (π.χ. USB), οι πραγματικές διαδρομές δεδομένων και οι μορφές είναι εντελώς ξεχωριστές.

Εν συντομία, το MIDI μεταδίδει πληροφορίες που καθοδηγούν ένα όργανο να παίζει ή μια συσκευή να εκτελέσει μια λειτουργία. Σε αντιστοιχία με ένα μουσικό κουτί, ένα αρχείο MIDI ή μια ροή δεδομένων είναι απλώς ένα σύνολο οδηγιών που μεταδίδονται μέσω ενός καλωδίου με σειριακό τρόπο, αλλά όταν ένα ηλεκτρονικό όργανο ερμηνεύει τα δεδομένα αρχίζουμε να ακούμε ήχο.

Πριν την εφαρμογή του MIDI με την εξαίρεση κάποιων περιπτώσεων ήταν αναγκαίο ένα κομμάτι να εκτελείται σε πραγματικό χρόνο. Με την εισαγωγή της ηλεκτρονικής παραγωγής μουσικής και του MIDI, μια μουσική εκτέλεση μπορεί να καταγραφεί στον ψηφιακό τομέα και στη συνέχεια να αναπαραχθεί πιστά σε ένα περιβάλλον παραγωγής που μιμείται την παραδοσιακή μορφή και τις λειτουργίες της πολυκάναλης εγγραφής. Τα βασικά μέρη θα μπορούσαν να εγγράφονται ένα κάθε φορά, επιτρέποντας τη δημιουργία μιας σύνθεσης χρησιμοποιώντας διάφορα ηλεκτρονικά όργανα.

Το MIDI επέτρεψε επιτέλους την επεξεργασία, την επίστρωση, την αλλαγή και τη βελτίωση ενός μουσικού κομματιού με σχετική ευκολία και υπό πλήρως αυτοματοποιημένο έλεγχο υπολογιστή. Η δυνατότητα για μελλοντική επέκταση και ο αυξημένος έλεγχος σε ένα ολοκληρωμένο σύστημα παραγωγής έχει ωθήσει την ανάπτυξη μιας βιομηχανίας που είναι επιπλέον πολύ προσιτή. Για πρώτη φορά στην ιστορία της μουσικής, είναι δυνατό για ένα άτομο να πραγματοποιήσει οικονομικά και στον κατάλληλο χρόνο για τον ίδιο μια μουσική παραγωγή πλήρους κλίμακας.

Με την εισαγωγή διάφορων ισχυρών εργαλείων υλικού ή λογισμικού, δεν είναι μόνο δυνατό αλλά και σχετικά εύκολο να δημιουργηθεί μια σύνθεση με τη χρήση ήχων που μιμούνται σχεδόν οποιοδήποτε όργανο που μπορεί να φανταστεί κανείς. Αρχικά, πολλοί μουσικοί συνοδείας μίλησαν εναντίον του MIDI, λέγοντας θα χρησιμοποιηθεί ως ρομπότ που θα τους αντικαθιστούσε, κάτι που δεν επαληθεύτηκε σε μεγάλο βαθμό, μια και πολλοί από αυτούς χρησιμοποιούν τώρα τη δύναμη του MIDI για να επεκτείνουν τις μουσικές τους ικανότητες και να δημιουργήσουν δικές τους παραγωγές. Σήμερα, το MIDI χρησιμοποιείται από πολλούς επαγγελματίες και μη επαγγελματίες μουσικούς για να εκτελεί ένα εκτεταμένο φάσμα εργασιών παραγωγής, συμπεριλαμβανομένης της μουσικής παραγωγής και της ηχητικής κάλυψης βίντεο και ταινίας.

3.2 Τρόπος λειτουργίας

MIDI Message

Το MIDI επικοινωνεί ψηφιακά δεδομένα μουσικής απόδοσης μεταξύ συσκευών ως σειρά μηνυμάτων MIDI. Αυτά τα μηνύματα μεταδίδονται παραδοσιακά μέσω μιας τυπικής γραμμής MIDI με σειριακό τρόπο με ταχύτητα 31.250 bit / sec. Τα μηνύματα MIDI αποτελούνται από ομάδες λέξεων 8-bit (γνωστές ως byte), οι οποίες μεταδίδονται με σειριακό τρόπο για να μεταφέρουν μια σειρά οδηγιών σε μία ή όλες τις συσκευές MIDI μέσα σε ένα σύστημα.

Μόνο δύο τύποι byte καθορίζονται από την προδιαγραφή MIDI: το status byte και το data byte.

- Ένα status byte χρησιμοποιείται για τον προσδιορισμό του τύπου λειτουργίας MIDI που θα εκτελεστεί από μια συσκευή ή πρόγραμμα. Χρησιμοποιείται επίσης για την κωδικοποίηση δεδομένων καναλιού (επιτρέποντας τη λήψη της εντολής από μια συσκευή που έχει ρυθμιστεί να ανταποκρίνεται στο επιλεγμένο κανάλι).
- Ένα data byte χρησιμοποιείται για να συσχετίσει μια τιμή με το συμβάν που δίνεται από το συνοδευτικό status byte.

Αν και ένα byte αποτελείται από οκτώ bit, το πιο σημαντικό bit (MSB, το αριστερότερο δυαδικό bit σε μια ψηφιακή λέξη) χρησιμοποιείται αποκλειστικά για τον προσδιορισμό του τύπου του byte. Το MSB ενός status byte είναι πάντα 1, ενώ το MSB ενός data byte είναι πάντα 0. Για παράδειγμα, το μήνυμα του σχήματος 3.2 θα μεταδώσει οδηγίες που θα διαβάζονται ως εξής: "Μετάδοση μηνύματος Note-On μέσω του καναλιού MIDI # 5, χρησιμοποιώντας την νότα # 64, με επίθεση ταχύτητας (επίπεδο έντασης μιας νότας) 89".[6]

	<u>Status Byte</u>	<u>Data Byte 1</u>	<u>Data Byte 2</u>
Περιγραφή:	Status/channel #	Note#	Attack velocity
Δυαδικά δεδομένα:	(1001.0100)	(0100.0000)	(0101.1001)
Αριθμητική τιμή:	(Note on/CH#5)	(64)	(89)

Σχήμα 3.2: Παράδειγμα μηνύματος MIDI

MIDI Channels

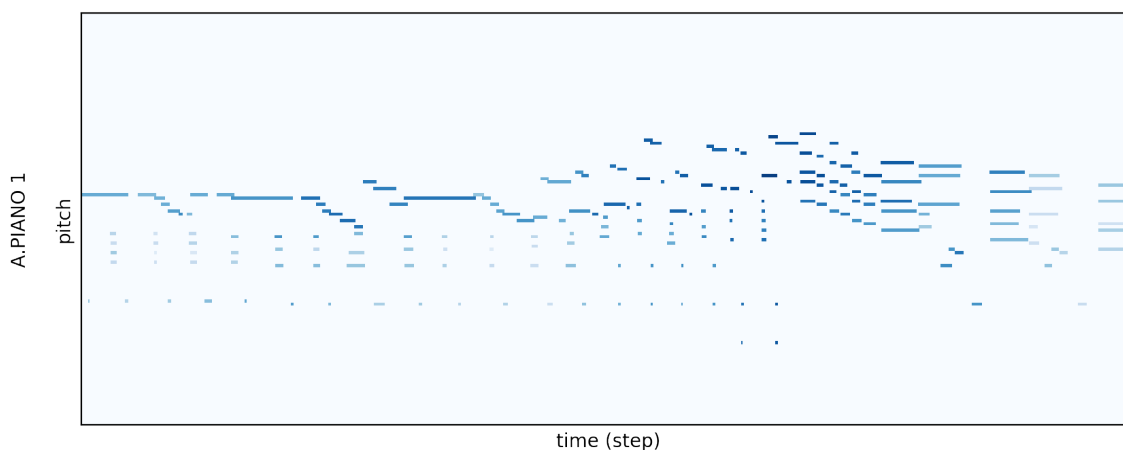
Ακριβώς όπως ένας δημόσιος ομιλητής μπορεί να ξεχωρίσει ένα άτομο σε πλήθος και να του μεταδώσει ένα ή περισσότερα μηνύματά, τα μηνύματα MIDI μπορούν να κατευθυνθούν για να μεταδώσουν πληροφορίες σε μια συγκεκριμένη συσκευή ή μια σειρά συσκευών μέσα σε ένα σύστημα MIDI. Αυτό γίνεται με την ενσωμάτωση μιας κατάληξης 4 bit που σχετίζονται με το κανάλι εντός του status / αριθμού καναλιού byte δίνοντάς του τη μορφή (1SSS CCCC). Αυτό καθιστά δυνατή την μετάδοση πληροφοριών απόδοσης ή ελέγχου σε μια συγκεκριμένη συσκευή - ή μια γεννήτρια ήχου μέσα σε μια συσκευή - που έχει εκχωρηθεί σε ένα συγκεκριμένο κανάλι.

Δεδομένου ότι γίνεται χρήση 4 bit, μπορούν να μεταδοθούν έως και 16 διακριτά κανάλια MIDI μέσω ενός μόνο καλωδίου MIDI ή καθορισμένης θύρας.

Κάθε φορά που μια συσκευή MIDI, μια γεννήτρια ήχου μέσα σε μια συσκευή ή μια λειτουργία προγράμματος έχει εντολή να ανταποκρίνεται σε έναν συγκεκριμένο αριθμό καναλιού, θα ανταποκρίνεται μόνο σε μηνύματα που μεταδίδονται σε αυτό το κανάλι (δηλαδή, αγνοεί τα μηνύματα καναλιών που μεταδίδονται σε οποιοδήποτε άλλο κανάλι).

3.3 Pianoroll

Το Pianoroll είναι μια μορφή αποθήκευσης μουσικής, που αντιπροσωπεύει ένα μουσικό κομμάτι σε έναν πίνακα σαν παρτιτούρα: Οι κατακόρυφοι και οριζόντιοι άξονες αντιπροσωπεύουν αντίστοιχα το pitch τής νότας και τον χρόνο. Οι τιμές αντιπροσωπεύουν την ταχύτητα (ενελοσιτψ) μιας νότας. Στην περίπτωση χρήσης πολλαπλών οργάνων σε ένα μουσικό κομμάτι σε μορφή Pianoroll, το Pianoroll χρησιμοποιεί αντίστοιχο πλήθος ξεχωριστών καναλιών (tracks). Τα μουσικά κομμάτια που χρησιμοποιήθηκαν ως δεδομένα για τη διενέργεια των πειραμάτων σε αυτήν την εργασία αποτελούσαν αρχεία τύπου Pianoroll στην αρχική τους μορφή.



Σχήμα 3.3: Παράδειγμα pianoroll

Αν και ένα αρχείο pianoroll περιέχει όλες τις απαραίτητες πληροφορίες, η απαραίτητη προεργασία για τον σχεδιασμό των τελικών δεδομένων που θα χρησιμοποιηθούν στο συγκεκριμένο εγχείρημα καθιστά απαραίτητη την μετατροπή του Pianoroll σε MIDI μέσω της βιβλιοθήκης python `pyrianoroll` [7] επειδή η πιο διαδεδομένη χρήση του MIDI προσφέρει περισσότερα εργαλεία χειρισμού των παραμέτρων με μεγαλύτερη αναλυτικότητα.

4.1 Προσέγγιση με βάση τη δομή του MIDI και τον ορισμό του στυλ

Όπως προαναφέρθηκε, οι πληροφορίες που περιέχονται σε ένα αρχείο MIDI είναι πιο κοντά σε μια αναπαράσταση παρτιτούρας ενός κομματιού από ένα αρχείο ήχου. Αποτελούν οδηγίες σαφώς ορισμένες, οι οποίες όμως μπορούν να εκληφθούν με διάφορους τρόπους ανάλογα με το μέσο που θα επιλεγεί να τις αποδώσει. Επομένως, ένα από τα κύρια χαρακτηριστικά το οποίο δεν εμπεριέχεται εξαρχής σε MIDI αρχεία είναι το ηχόχρωμα (χροιά - timbre) των διαφόρων ήχων που απαρτίζουν ένα κομμάτι. Το ηχόχρωμα παρέχει πολλές αντιληπτικές ενδείξεις για την κατηγοριοποίηση και τον προσδιορισμό των πηγών ήχου[8]. Έχει τη δυνατότητα να επηρεάζει μεμονωμένα την συναισθηματική αντίληψη της μουσικής ανεξαρτήτως άλλων ακουστικών, νοητικών και παραστατικών παραγόντων[9] και είναι ένα από τα κύρια μέσα αναγνώρισης, ταυτοποίησης και παρακολούθησης κατά την πάροδο του χρόνου μιας πηγής ήχου και συνεπώς εμπλέκεται στην απόλυτη κατηγοριοποίηση ενός ηχητικού αντικειμένου[10]. Κατ' επέκταση λοιπόν μπορεί να εξαχθεί το συμπέρασμα ότι είναι υψίστης σημασίας στην αναγνώριση του στυλ ή ακόμα και του γενικότερου είδους ενός μουσικού κομματιού από έναν ακροατή.

Η έννοια της χροιάς προσεγγίζεται σε έναν βαθμό στα αρχεία MIDI με specification GM 1 (General MIDI Level 1) όπου κάθε channel αντιστοιχεί σε μια προτεινομένη ομάδα οργάνων η οποία μπορεί να εφαρμοστεί χειρωνακτικά για να αποδοθεί το κομμάτι με μεγαλύτερη πιστότητα σε μια ζωντανή εκτέλεση. Η πληροφορία του οργάνου που χρησιμοποιείται, αν και ιδιαίτερα χρήσιμη για προσδιορισμό και μεταφορά είδους/στυλ, κάνει σημαντικά πιο πολύπλοκη την κωδικοποίηση πληροφοριών σε μορφή επαρκή για την εκπαίδευση ενός νευρωνικού δικτύου. Αγνοώντας λοιπόν την μία θεωρητική ένδειξη χροιάς αυτού του τύπου αρχείων για πρακτικούς λόγους, επιχειρείται μετατροπή των κομματιών καθαρά σε επίπεδο σύνθεσης.

Τα αποτελέσματα χρήσης συμβολικών δεδομένων όπως το MIDI για την αναπαράσταση μουσικών πληροφοριών στο συναφές εγχείρημα κατηγοριοποίησης κομματιών σε μουσικά είδη, υποδεικνύουν επάρκεια αυτού του επιπέδου ανάλυσης σε ορισμένες περιπτώσεις, αφήνοντας ανοιχτές τις προοπτικές της παρούσας προσέγγισης.

Μέρος II
Πρακτικό Μέρος

Οργάνωση δεδομένων

Για την εκτέλεση τού πειράματος χρησιμοποιήθηκε το Lakh Pianoroll Dataset (LPD)[11][12] για την συλλογή των pianoroll αρχείων των μουσικών κομματιών και την επεξεργασία τους. Στην συνέχεια, τα μουσικά κομμάτια κωδικοποιήθηκαν με χρήση της βιβλιοθήκης Pypianoroll της Python, ώστε να έρθουν σε μορφή επεξεργάσιμη, κατάλληλη για την εξαγωγή των αποτελεσμάτων που επιδιώκει το πείραμα.

5.1 Lakh labels

Το LPD αποτελεί συλλογή από multitrack pianoroll, μέρος του οποίου έχει αντιστοιχηθεί με το Million Song Dataset (MSD) κάνοντας έτσι εφικτή την χρήση διαφόρων metadata.

Η εκδοχή που χρησιμοποιήθηκε είναι η lpd-17-cleansed η οποία περιέχει «21,425 seventeen-track pianorolls» με συγκεκριμένα χαρακτηριστικά τα οποία είναι σημαντικής χρησιμότητας για το στάδιο της εκπαίδευσης. Συγκεκριμένα, σε σχέση με το γενικό LPD:

- Έχουν αφαιρεθεί τα κομμάτια με πάνω από ένα time signature change event
- Έχουν αφαιρεθεί τα κομμάτια με ρυθμό διαφορετικό από 4/4
- Έχουν αφαιρεθεί τα κομμάτια των οποίων η πρώτη νότα δεν ξεκινάει από το χρόνο μηδέν
- Έχει κρατηθεί μόνο το αρχείο με το μεγαλύτερο βαθμό εμπιστοσύνης στην αντιστοίχιση για κάθε κομμάτι

Επιπλέον, τα pianoroll tracks έχουν συγχωνευτεί σε drums και δεκαέξι οικογένειες οργάνων σύμφωνα με τους αριθμούς προγράμματος που παρέχονται στα αρχεία MIDI και τις προδιαγραφές του General MIDI Level 1 (GM 1). Τα δεκαεφτά tracks είναι τα εξής:

<i>Drums</i>	<i>Guitar</i>	<i>Brass</i>	<i>Synth Pad</i>
<i>Piano</i>	<i>Bass</i>	<i>Reed</i>	<i>Synth Effects</i>
<i>Chromatic Percussion</i>	<i>Strings</i>	<i>Pipe</i>	<i>Ethnic</i>
<i>Organ</i>	<i>Ensemble</i>	<i>Synth Lead</i>	<i>Percussive</i>
			<i>Sound Effects</i>

Τέλος, στο lpd-17-cleansed κάθε κομμάτι έχει αντιστοίχιση στο MSD. Η ιδιότητα αυτή είναι κομβικής σημασίας καθώς για να επιχειρηθεί ένας σχεδιασμός μοντέλου αλλαγής στυλ είναι απαραίτητη η εκπαίδευση του πάνω σε επαρκή δεδομένα καθορισμένων κατηγοριών. Συγκεκριμένα, γίνεται χρήση των Last.fm labels τα οποία αποδίδονται στο μεγαλύτερο υποσύνολο του MSD και τα οποία περιέχουν 69 χαρακτηρισμούς κομματιών από τη βάση δεδομένων του Last.fm και τα κομμάτια του MSD στα οποία έχουν αποδοθεί. Οι χαρακτηρισμοί κυμαίνονται από επίσημους όρους μουσικών ειδών (rock, jazz, funk) σε πιο αφηρημένες περιγραφές διάθεσης (happy, melancholy, relax). Τα labels που χρησιμοποιούνται είναι προσαρμοσμένα στο LPD και έτσι υπάρχει μια καλή εκτίμηση για την οργάνωση του dataset ανά είδος και στυλ.

5.2 Κωδικοποίηση πληροφοριών

Μετά την οργάνωση των αρχείων ανά είδος σύμφωνα με τα Last.fm labels ακολούθησε η μετατροπή σε χρηστική μορφή για το εγχείρημα της εργασίας. Αρχικά εξήχθησαν τα κομμάτια σε αρχεία τύπου .mid από τη μορφή αποθήκευσης .npz, στην οποία βρίσκονται τα δεδομένα που περιέχει το LPD, με χρήση της Python βιβλιοθήκης PyPianoRoll. Τα .npz πρακτικά είναι αρχεία zip τα οποία περιέχουν:

- component arrays των piano rolls σε συμπιεσμένο sparse column format:
 - pianoroll_[index]_csc_data.npy
 - pianoroll_[index]_csc_indices.npy
 - pianoroll_[index]_csc_indptr.npy
- tempo.npy: τον πίνακα tempo
- downbeat.npy: τον πίνακα down beat
- info.json: ένα αρχείο JSON το οποίο περιέχει meta data και πληροφορίες του κομματιού

Στη συνέχεια έγινε προεπεξεργασία των MIDI αρχείων έτσι ώστε να αναχθούν σε μορφή κατάλληλη για εκπαίδευση από πλευράς συμβατότητας αλλά και απλότητας. Για την εκπλήρωση του σκοπού αυτού έγινε εκτενής χρήση του music21 toolkit.

Καθώς το εγχείρημα αποσκοπεί στην αλλαγή είδους ολόκληρου κομματιού, δηλαδή στο αρμονικό σύνολο που δημιουργείται από όλα τα όργανα που χρησιμοποιούνται και λόγω του τρόπου με τον οποίο αναπαρίστανται οι πληροφορίες σε ένα MIDI αρχείο, θα ήταν αρκετά πολύπλοκη η σύμπτυξη των κρουστών με τα υπόλοιπα όργανα σε μία διανυσματική αναπαράσταση. Για αυτόν τον λόγο έγινε απαλοιφή των track drums και percussive του κάθε αρχείου.

Επιπλέον περιορίστηκαν οι πληροφορίες που είναι διαθέσιμες στα αρχεία του LPD στις απολύτως απαραίτητες για την επαναφορά ενός φαινομενικά ίδιου σε αρμονικό επίπεδο αρχείου MIDI. Με προτεραιότητα την όσο το δυνατόν πιο συνοπτική, αλλά και επαρκή, αναπαράσταση

της μελωδίας του κομματιού, αποφασίστηκε η χρήση διάνυσμάτων για την περιγραφή κάθε μεμονωμένης νότας με τρία στοιχεία:

- το χρονικό σημείο στο οποίο προκύπτει
- την διάρκεια της
- τη συχνότητά της στα πλαίσια των 128 τιμών που παρέχει το GM 1

Η μεταβλητή του tempo αγνοήθηκε επίσης καθώς, αν και προφανής για τον ακροατή, θεωρήθηκε απλή η παραποίησή της χειρωνακτικά έτσι ώστε να βγάξουν νόημα ηχητικά τα αποτελέσματα. Τέλος, αγνοήθηκαν και τα υπόλοιπα χαρακτηριστικά ενός instance νότας που παρέχει το GM 1 (π.χ. velocity) καθώς δεν είναι απαραίτητα στο πλαίσιο μιας ανάλυσης σύνθεσης ενός κομματιού MIDI.

Εν τέλει, το διάνυσμα περιγραφής νοτών μπόρεσε να περιγραφεί με τη χρήση των παραμέτρων του music21 ως εξής:

(offset, quarterlength, pitch)

όπου οι πρώτες δύο μεταβλητές κάνουν χρήση της διακριτής αναπαράστασης του χρόνου quarterlength που παρέχει το music21, η οποία χρησιμοποιεί σαν βάση μέτρησης για τη διάρκεια τα τέταρτα νότας και εκφράζει τα μεγέθη ως πολλαπλάσιά τους. Η μεταβλητή pitch μπορεί να πάρει τις ακεραίες τιμές από το 0 έως το 127.

Δοκιμάστηκαν διάφοροι τρόποι φόρτωσης αυτών των τριών τιμών στο σχεδιαζόμενο μοντέλο ώστε να ελεγχθεί η αντίστοιχη απόδοση του.

Στην αρχική και ενδεχομένως πιο ελεύθερη έκφανση, χρησιμοποιήθηκαν οι νότες κάθε κομματιού στο σύνολό τους, με χειρισμό των μεταβλητών offset και quarterlength ως συνεχόμενων δεκαδικών ανοιγμένων στο [0,1] όπως είθισται για συνεχόμενες τιμές στα πλαίσια της μηχανικής μάθησης. Για την μεταβλητή pitch, η διακριτότητα και το περιορισμένο πεδίο τιμών της προσέφερε τη δυνατότητα μετατροπής των τιμών σε One Hot Encoding (OHE) κωδικοποίηση. Κάθε τιμή μετατρέπεται σε διάνυσμα με μήκος 128, γεμάτο μηδενικά, εκτός από το 1 στη στήλη για τον ακεραίο που αντιπροσωπεύει το μοτίβο. Αυτό χρησιμεύει ώστε να μπορούμε να διαμορφώσουμε το δίκτυο για να προβλέψουμε την πιθανότητα καθενός από τους 128 διαφορετικούς αριθμούς στο πεδίο τιμών (μια ευκολότερη αναπαράσταση) αντί να προσπαθούμε να το αναγκάσουμε να προβλέψει ακριβώς την επόμενη τιμή[13].

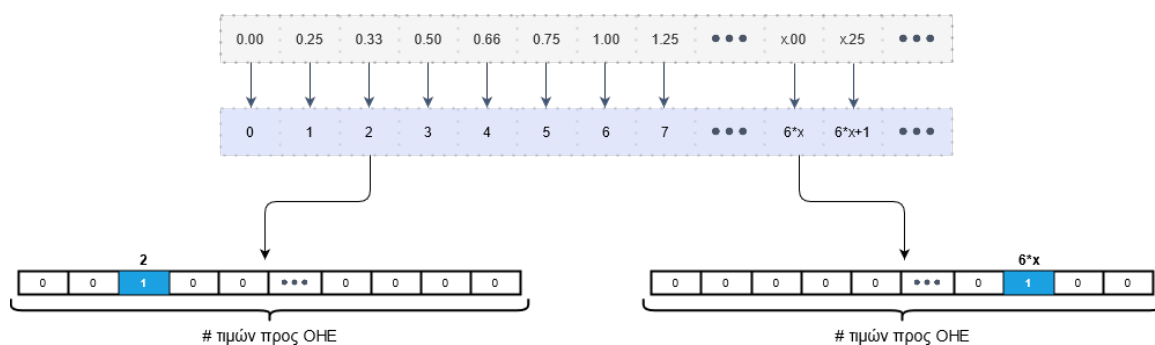
Καθώς οι συνεχόμενες τιμές είναι σημαντικά δυσκολότερες στον χειρισμό, τόσο ως είσοδος για το πρόγραμμα αλλά και ως έξοδος με ικανοποιητική ακρίβεια, ήταν σημαντική η εύρεση μιας διακριτοποίησης των μεταβλητών offset, quarterlength ώστε να δεχθούν αντίστοιχη κωδικοποίηση με το pitch. Δεδομένου του ότι τα πολλαπλάσια του quarterlength που χαρακτηρίζουν τις δύο μεταβλητές χρησιμοποιούν συγκεκριμένες υποδιαίρεσεις της μονάδας ήταν επιθυμητή η αντιστοίχσή τους σε πεπερασμένο πλήθος ακεραίων. Συγκεκριμένα, κάθε πολλαπλάσιο quarterlength παρουσιάζει δεκαδικό μέρος 00, 25, 33, 50, 66 ή 75. Μπορεί λοιπόν το πεδίο τιμών offset, quarterlength να αναχθεί σε αντίστοιχο ακεραίων με εξαπλάσιο μέγιστο πάνω στο οποίο μπορεί να εφαρμοστεί OHE (Σχήμα 5.1).

Σε αυτήν την περίπτωση είναι αναγκαίος ο περιορισμός των τιμών που μπορούν να πάρουν offset και quarterlength για λόγους χωρητικότητας και ταχύτητας. Σε διάφορα dataset συγκεκριμένων ειδών οι μέγιστες τιμές τους ξεπερνούν το 1000.0 και το 600.0 αντίστοιχα, καθιστώντας δύσκολο έως αδύνατο τον εξαπλασιασμό των πεδίων τιμών. Παρ'όλα αυτά, σε έλεγχο κατανομής τιμών για αυτές τις μεταβλητές διαπιστώθηκε πως πάνω από το 95% των δεδομένων κυμαίνονταν σε σημαντικά χαμηλότερες τιμές. Κατά την κωδικοποίηση αυτού του τύπου λοιπόν ορίστηκαν ως μέγιστα το 600.0 για το offset και το 50.0 για το quarterlength χωρίς ιδιαίτερο συμβιβασμό από πλευράς απώλειας πληροφοριών και διατηρώντας ένα σύνολο αρμονικά άρτιων κομματιών (νότες με διάρκεια άνω των 50 quarterlength ή και ακόμα χαμηλότερα δεν ανταποκρίνονται σε συνηθισμένη μουσική επιλογή ως επί το πλείστον).

Με στόχο την απόκτηση πληροφοριών σχετικά με τις αρμονικές συνήθειες κάθε είδους, έγινε προσπάθεια μεταφοράς της μεταβλητής pitch σε κοινή βάση για όλα τα κομμάτια. Συγκεκριμένα, θεωρήθηκε πως με την μετατόπιση της τονικής των κομματιών σε μία καθολική νότα για όλα τα είδη, αυξανόταν η πιθανότητα αναγνώρισης διαφόρων μουσικών μοτίβων (π.χ. κλίμακα) ενός είδους. Με αυτόν τον τρόπο, στα πλαίσια της rock για παράδειγμα, η πεντατονική κλίμακα της Ντο δεν χρειάζεται να αντιμετωπιστεί ως διαφορετική από την πεντατονική της Λα, ωθώντας σε ένα καλό θεωρητικό σενάριο στην αναγνώριση της χρήσης πεντατονικής ως χαρακτηριστικό γνώρισμα του είδους.

Για το εγχείρημα αυτό χρησιμοποιήθηκε το music21 για αναγνώριση και μεταφορά τονικότητας. Μέσω του toolkit εφαρμόστηκε ο αλγόριθμος Krumhansl-Schmuckler, μια από τις πιο γνωστές αυτοματοποιημένες μεθόδους εύρεσης τονικότητας και κλίμακας, για να γίνει το κατάλληλο transpose που θα φέρει το κάθε κομμάτι σε τονική Ντο (C). Η σχετικά απρόβλεπτη συμπεριφορά του αλγορίθμου που έχει παρατηρηθεί σε κάποια κομμάτια δεν τον καθιστά ιδανικό για είδη πιο πολύπλοκα σε επίπεδο αρμονίας, η μέθοδος που ακολουθεί όμως κατά πάσα πιθανότητα είναι ικανοποιητική για την αναγωγή κομματιών σε τουλάχιστον αντίστοιχη αρμονική βάση[14]. Επιπλέον δυσκολία σε αυτό το στάδιο μπορεί να παρουσιαστεί εάν εμφανίζεται αλλαγή τονικής ή κλίμακας εντός ενός κομματιού, κάτι το οποίο δεν παρέχεται από το dataset ως διευκρίνηση.

Σε μια καθαρά ρυθμική προσέγγιση της έννοιας του στυλ, κρατήθηκαν μόνο οι μεταβλητές offset και quarterlength αγνοώντας το κομμάτι της αρμονίας. Στο παρελθόν, έχει αναφερ-

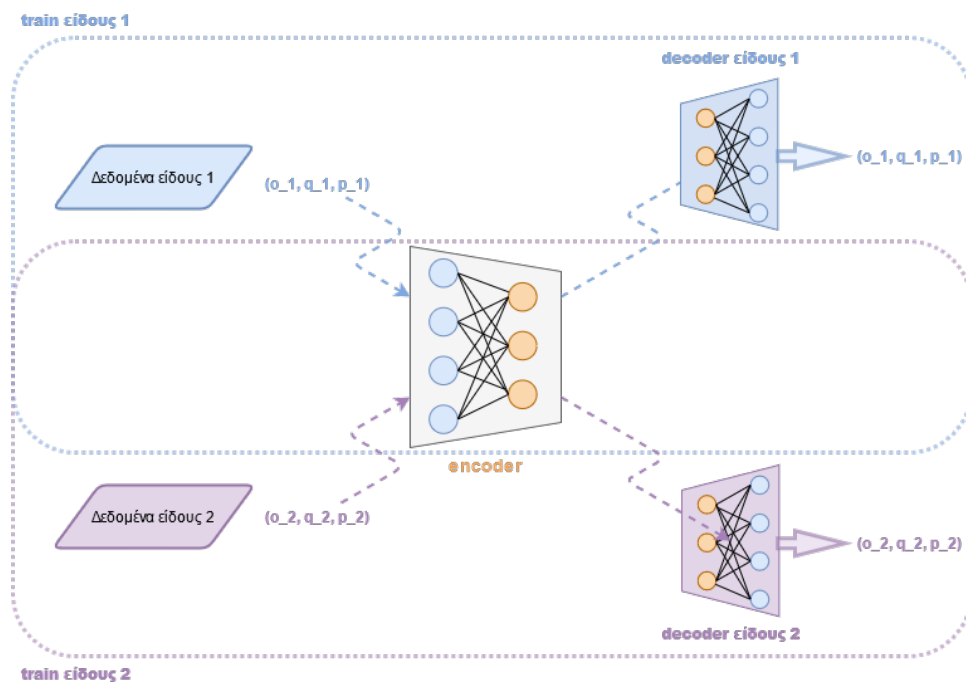


Σχήμα 5.1: Αναπαράσταση τιμών quarterlength σε OHE πίνακες μετά από αντιστοίχιση σε ακεραίους

θεί σε αρκετές περιπτώσεις η σημασία του ρυθμού σε εγχειρήματα κατηγοριοποίησης και σε ορισμένες περιπτώσεις έχει τεθεί ως επίκεντρο της έρευνας[15][16]. Με απαλοιφή του pitch όσον αφορά την είσοδο του μοντέλου επιχειρείται να εντοπιστούν ρυθμικά μοτίβα που πιθανόν χαρακτηρίζουν επαρκώς ορισμένα είδη μουσικής.

6.1 Autoencoder με πολλαπλούς decoder και κοινό encoder

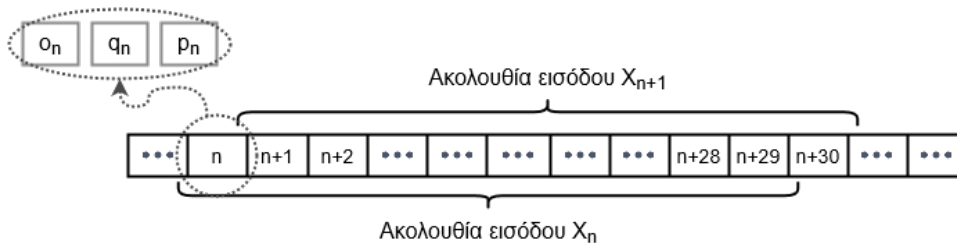
Το μοντέλο βασίστηκε στην αρχιτεκτονική που σχεδιάστηκε από τους Jesse Engel et al. στην εργασία “Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders” [17]. Πρόκειται για σχεδιασμό autoencoder με χρήση ξεχωριστών decoder για κάθε είδος και ενός καθολικού encoder. Η αρχιτεκτονική αυτή έχει εφαρμοστεί για το ίδιο εγχείρημα στην εργασία “A Universal Music Translation Network” όπου γίνεται χρήση αρχείων audio για αναπαράσταση της μουσικής[18]. Λόγω της χρήσης MIDI αναπαραστάσεων στην παρούσα εργασία δεν γίνεται χρήση του Wavenet πρωτοκόλλου σε αντίθεση με τις παραπάνω έρευνες.



Σχήμα 6.1: Γραφική απεικόνιση του μοντέλου

Το γενικό μοντέλο απαρτίζεται από έναν αριθμό επιμέρους μοντέλων ξεχωριστά ορισμένων, τα οποία αλληλεπιδρούν μεταξύ τους και μοιράζονται κοινά στοιχεία. Για την εκπαίδευση ορίζεται ο κοινός encoder ο οποίος σε συνδυασμό με τον κάθε decoder ξεχωριστά σχηματίζει το αντίστοιχο train μοντέλο που προορίζεται για τον χειρισμό πληροφοριών ενός είδους ως autoencoder. Παράλληλα, έχουν οριστεί για τον encoder και τους decoder ανάλογα inference μοντέλα για την διεξαγωγή προβλέψεων με τα βάρη που προκύπτουν από τις εκπαιδεύσεις. Τα μοντέλα inference έχουν κοινή γενική δομή με τα μοντέλα που παραλληλίζουν, με τροποποιημένες μορφές εισόδων και εξόδων όπως απαιτείται για την διεξαγωγή προβλέψεων.

Οι εισοδοί και εξοδοί που δίνονται στον autoencoder αποτελούν ακολουθίες 30 στοιχείων. Κάθε instance των ακολουθιών είναι ένα διάνυσμα (offset, quarterlength, pitch) που όπως προαναφέρθηκε, είναι η επιλεχθείσα αναπαράσταση μιας νότας και των απαραίτητων χαρακτηριστικών της. Το μήκος των ακολουθιών περιορίστηκε σε 30 στοιχεία καθώς θεωρήθηκε ισορροπημένη επιλογή ενός μεγέθους αρκετά μικρού ώστε να είναι διαχειρίσιμο για το μοντέλο κατά τη διαδικασία εκμάθησης και αρκετά μεγάλου ώστε να εμπεριέχονται πληροφορίες μουσικά επαρκείς για τον ορισμό ενός είδους.



Σχήμα 6.2: Γραφική απεικόνιση των ακολουθιών

Οι ακολουθίες σχηματίζονται ορίζοντας σειριακά ως αρχή τους κάθε νότα του dataset και συμπληρώνοντας τες με τις επόμενες 29. Τα κομμάτια έχουν περαστεί στο dataset διατηρώντας τη δομή τους οπότε κάθε ακολουθία αποτελεί μια μουσική φράση 30 νοτών.

Στα πλαίσια της εργασίας ο αριθμός των decoder περιορίζεται σε δύο κάνοντας εφικτή την μεταφορά είδους ανά συγκεκριμένο ζεύγος τη φορά. Μεγαλύτερος αριθμός δεν είναι πρακτικός σε αρχικό στάδιο χωρίς αξιολόγηση κάποιων αποτελεσμάτων λόγω των πόρων και του χρόνου που θα απαιτούσε ένα τέτοιο μοντέλο συγκριτικά.

6.2 Seq2seq

Το μοντέλο ακολουθεί το πρωτόκολλο ενός συνηθισμένου encoder-decoder model το οποίο αποτελεί τρόπο οργάνωσης επαναλαμβανόμενων νευρωνικών δικτύων (RNN) για χρήση σε προβλήματα πρόβλεψης sequence-to-sequence.

Το Recurrent Neural Network (RNN) είναι μια φυσική γενίκευση των feedforward νευρωνικών δικτύων σε ακολουθίες. Δεδομένης μιας ακολουθίας εισόδων (x_1, \dots, x_T) , ένα τυπικό RNN υπολογίζει μια ακολουθία εξόδων (y_1, \dots, y_T) επαναλαμβάνοντας την ακόλουθη εξίσωση:

$$h_t = \text{sigm}(W^{hx}x_t + W^{hh}h_{t-1})$$

$$y_t = W^{yh}h_t$$

Το RNN μπορεί εύκολα να αντιστοιχίσει ακολουθία σε ακολουθία όποτε η ευθυγράμμιση μεταξύ των εισόδων και των εξόδων είναι γνωστή εκ των προτέρων. Στη συγκεκριμένη περίπτωση, κάθε ζεύγος encoder-decoder που αντιστοιχεί σε συγκεκριμένο είδος ακολουθεί λογική autoencoder όπου η έξοδος ταυτίζεται με την είσοδο. Με αυτήν την πρακτική, στο ιδανικό σενάριο ένα κομμάτι που περνάει από τον decoder του είδους στο οποίο ανήκει δεν δέχεται μεταβολές. Περνώντας από decoder άλλου είδους παίρνει χαρακτηριστικά τα οποία το μοντέλο έχει διακρίνει ως αντιπροσωπευτικά του είδους.

Η απλούστερη στρατηγική για τη γενική εκμάθηση ακολουθίας είναι η αντιστοίχιση της input sequence σε ένα σταθερού μεγέθους διάνυσμα χρησιμοποιώντας ένα RNN και στη συνέχεια η χαρτογράφηση του διανύσματος στην target sequence με ένα άλλο RNN. Παρόλο που κάτι τέτοιο θα μπορούσε θεωρητικά να λειτουργήσει αφού στο RNN παρέχονται όλες οι σχετικές πληροφορίες, θα ήταν δύσκολο να εκπαιδευτούν τα RNN λόγω των μακροπρόθεσμων εξαρτήσεων που προκύπτουν (εικόνα 1). Ωστόσο, ένα Long Short-Term Memory (LSTM) [16] είναι γνωστό ότι μαθαίνει προβλήματα με χρονικές εξαρτήσεις μεγάλης εμβέλειας, οπότε μπορεί να αποφέρει αποτέλεσμα σε αυτές τις συνθήκες.

Ο στόχος του LSTM είναι να εκτιμηθεί η εξαρτημένη πιθανότητα $p(y_1, \dots, y_{T'} | x_1, \dots, x_T)$ όπου (x_1, \dots, x_T) είναι μια ακολουθία εισόδου και $(y_1, \dots, y_{T'})$ είναι η αντίστοιχη ακολουθία εξόδου της οποίας το μήκος T' μπορεί να διαφέρει από το T . Το LSTM υπολογίζει αυτήν την εξαρτημένη πιθανότητα λαμβάνοντας πρώτα την σταθερών διαστάσεων αναπαράσταση v της ακολουθίας εισόδου (x_1, \dots, x_T) που δίνεται από το τελευταίο hidden state του LSTM και, στη συνέχεια, υπολογίζοντας την πιθανότητα του $(y_1, \dots, y_{T'})$ με μια τυπική διαμόρφωση LSTM-LM της οποίας το αρχικό hidden state έχει οριστεί στην αναπαράσταση v του (x_1, \dots, x_T) :

$$p(y_1, \dots, y_{T'} | x_1, \dots, x_T) = \prod_{t=1}^{T'} p(y_t | v, y_1, \dots, y_{t-1})$$

Σε αυτήν την εξίσωση, κάθε κατανομή $p(y_t | v, y_1, \dots, y_{t-1})$ αντιπροσωπεύεται με ένα softmax πάνω σε όλες τις ακολουθίες στο λεξιλόγιο[19].

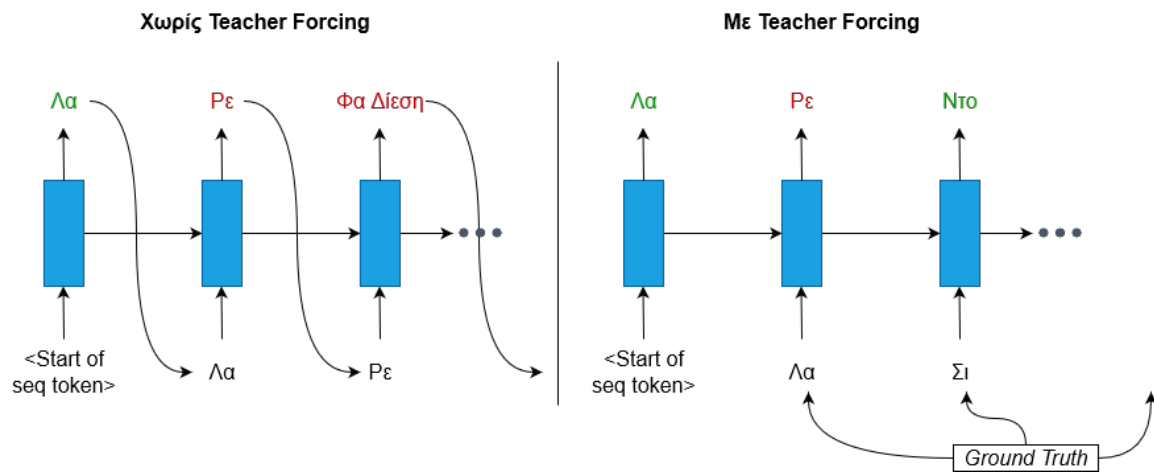
Βασικό στις προαναφερθείσες έρευνες το μοντέλο που σχεδιάστηκε διαφέρει από την παραπάνω περιγραφή ως εξής: χρησιμοποιεί δύο διαφορετικά LSTM, ένα για την ακολουθία εισόδου και ένα άλλο για την ακολουθία εξόδου και, καθώς τα deep LSTM ξεπερνούν σημαντικά τα shallow LSTM, επιλέγονται LSTM με πολλαπλά layers.

6.3 Teacher Forcing

Στα πλαίσια της προσέγγισης seq2seq γίνεται χρήση teacher forcing[20]. Κατά την εκπαίδευση μοντέλων πρόβλεψης ακολουθιών είναι σύνηθες να χρησιμοποιείται η έξοδος του προηγούμενου χρονικού βήματος $y(t)$ ως είσοδος του επόμενου βήματος $X(t+1)$. Με αυτή την τακτική όμως μια λανθασμένη πρόβλεψη εκτροχιάζει και την υπόλοιπη ακολουθία, παρουσιάζοντας έτσι προβλήματα όπως αργή σύγκλιση και αστάθεια.

Το teacher forcing είναι μία στρατηγική εκπαίδευσης RNN κατά την οποία η αναμενόμενη έξοδος $y(t)$ είναι αυτή που χρησιμοποιείται ως είσοδος του επόμενου βήματος $X(t+1)$ αντί

για την παραγόμενη από το μοντέλο έξοδο. Η προσέγγιση αυτή βελτιώνει τις ικανότητες και τη σταθερότητα ενός μοντέλου πρόβλεψης ακολουθιών.



Σχήμα 6.3: Στάδιο decoding με και χωρίς χρήση Teacher Forcing

Προετοιμασία και Υπερπαραμέτροι

Πριν από την εκκίνηση εκπαίδευσης του κάθε μοντέλου έγινε διαχωρισμός των εκάστοτε δεδομένων που θα χρησιμοποιηθούν σε train και test sets. Το train set αποτελείται από το 90% των συνολικών ακολουθιών που εξάγονται από τα δεδομένα πάνω στις οποίες γίνεται η εκπαίδευση, με το τελευταίο 10% περίπου να διατηρείται για μετέπειτα ελέγχους απόδοσης του μοντέλου (test set).

Η επιλογή κατάλληλων υπερ-παραμέτρων (hyperparameters) έγινε με γνώμονα τον τρόπο κωδικοποίησης των δεδομένων (συνεχόμενες τιμές ή OHE).

7.1 Συναρτήσεις ενεργοποίησης

Η επιλογή της συνάρτησης ενεργοποίησης ελέγχει πόσο καλά το μοντέλο μαθαίνει το σύνολο δεδομένων εκπαίδευσης και επιπλέον καθορίζει τον τύπο των προβλέψεων που μπορεί να κάνει το μοντέλο.

Καθώς οι έξοδοι αποτελούνται από τρεις διαφορετικές μεταβλητές που περνάνε από ξεχωριστά Dense Layers, ορίζεται συνάρτηση ενεργοποίησης ανάλογα με την μορφή αναπαράστασής τους. Συγκεκριμένα, για το μοντέλο στο οποίο τα offset, quarterlength αναπαρίστανται με συνεχόμενες τιμές μεταξύ 0 και 1 εφαρμόστηκε η ReLu[21] και Sigmoid[22] αντίστοιχα μετά από διάφορες δοκιμές συνδυασμού αυτών των δύο συναρτήσεων. Σε κάθε περίπτωση κωδικοποίησης OHE εφαρμόστηκε η Softmax όπως είδισται για αναπαραστάσεις μεταβλητής τέτοιου είδους[23].

7.2 Συναρτήσεις Κόστους

Κατά την εκπαίδευση, η απόκλιση της τρέχουσας κατάστασης του μοντέλου πρέπει να υπολογίζεται επαναλαμβανόμενα. Αυτό απαιτεί την επιλογή μιας συνάρτησης σφάλματος, καλούμενη κατά σύμβαση συνάρτηση κόστους (loss function), η οποία χρησιμοποιείται για να εκτιμηθούν οι απώλειες του μοντέλου έτσι ώστε να ανανεωθούν τα βάρη και να μειωθούν οι απώλειες των επόμενων εκτιμήσεων[24].

Αντίστοιχα με τις συναρτήσεις ενεργοποίησης, η επιλογή των κατάλληλων συναρτήσεων κόστους εξαρτήθηκε από την εκάστοτε αναπαράσταση των μεταβλητών. Στις περιπτώσεις συνεχόμενων τιμών χρησιμοποιήθηκε η μέση τετραγωνική απόκλιση (Mean Squared Error), η πιο συνηθισμένη από τις συναρτήσεις κόστους παλινδρόμησης κατά τις οποίες γίνεται πρόβλεψη πραγματικών τιμών ενός μεγέθους. Η συνάρτηση ορίζεται τυπικά ως:

$$MSE \text{ formula} = (1/n) * \sum (actual - forecast)^2$$

Όπου: n = αριθμός αντικειμένων, Σ = άθροισμα, Actual = αρχική ή παρατηρούμενη τιμή του y , Forecast = τιμή y από παλινδρόμηση.[25]

Για τις ΟΗΕ αναπαραστάσεις χρησιμοποιήθηκε η κατηγορηματική διασταυρούμενη εντροπία (Categorical Crossentropy) η οποία χρησιμοποιείται σε περιπτώσεις κατηγοριοποίησης. Ο τύπος της συνάρτησης είναι:

$$-\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C \mathbf{1}_{y_i \in C_c} \log p_{model}[y_i \in C_c]$$

Το διπλό άθροισμα είναι πάνω από τις παρατηρήσεις « i », του οποίου ο αριθμός είναι « N », και των κατηγοριών « c », του οποίου ο αριθμός είναι « C ». Ο όρος « $\mathbf{1}_{\{y_i \in C_c\}}$ » είναι η συνάρτηση ένδειξης της παρατήρησης « i » που ανήκει στην κατηγορία εν« c ». Το « $p_{model}[y_i \in C_c]$ » είναι η πιθανότητα που προβλέπεται από το μοντέλο για την « i »τη παρατήρηση να ανήκει στην κατηγορία εν« c ». Όταν υπάρχουν περισσότερες από δύο κατηγορίες, το νευρικό δίκτυο εξάγει ένα διάλυσμα πιθανότητας « C », το καθένα δίνει την πιθανότητα ότι η είσοδος δικτύου πρέπει να ταξινομηθεί ως ανήκει στην αντίστοιχη κατηγορία [26].

7.3 Βελτιστοποιητές

Ο αλγόριθμος βελτιστοποίησης καθορίζει πώς ένα νευρωνικό δίκτυο χρησιμοποιεί τη διαφορά μεταξύ των αποτελεσμάτων που λαμβάνει κατά την εκπαίδευση και των τιμών που γνωρίζει ότι είναι αληθείς, για να ρυθμίσει τα βάρη στους κόμβους έτσι ώστε να κατευθυνθεί το δίκτυο προς μια λύση. Χρησιμοποιήθηκε καθολικά ο αλγόριθμος Adam[27] ο οποίος εκτός από την αποθήκευση ενός εκθετικά μειούμενου μέσου όρου προηγούμενων τετραγωνικών βαθμίδων, διατηρεί επίσης έναν εκθετικά μειούμενο μέσο όρο των προηγούμενων βαθμίδων χωρίς επεξεργασία τους[28].

7.4 Μετρικές

Στις περιπτώσεις συνεχόμενων τιμών τα αποτελέσματα αξιολογήθηκαν σύμφωνα με την ελαχιστοποίηση της μετρικής meansquarederror ενώ για τις περιπτώσεις ΟΗΕ όπου οι προβλέψεις κινούνται σε συγκεκριμένο σύνολο έγινε χρήση της accuracy metric (αριθμός σωστών προβλέψεων/αριθμός δεδομένων) στην οποία συνέκλιναν τα μοντέλα.

Η επιλογή ειδών προς μεταφορά εξαρτήθηκε από τον περιορισμό του dataset αλλά και από τα πρακτικά χαρακτηριστικά που ενδεχομένως επηρεάζουν την δυσκολία ενός τέτοιου εγχειρήματος. Ορίστηκαν η Rock, Jazz και Electronic για τις ανά ζεύγος μεταφορές είδους με κριτήριο τα στοιχεία που χαρακτηρίζουν και διαφοροποιούν το κάθε είδος σε επίπεδο σύνθεσης. Ως προς την πολυπλοκότητα, η Rock γενικά παρουσιάζει τα πιο λιτά χαρακτηριστικά ρυθμικά και αρμονικά σε αντίθεση με την Jazz που καλύπτει μεγάλο αρμονικό εύρος ως είδος και μπορεί να είναι ρυθμικά πολύπλοκη, και την Electronic στην οποία δίνεται συχνά μεγαλύτερη έμφαση στο ρυθμικό μέρος από πλευράς σύνθεσης αλλά κάνει χρήση πιο απλών αρμονικών κανόνων και παρουσιάζει επαναληπτικότητα στα μελωδικά μοτίβα. Δημιουργείται έτσι μία διαβάθμιση των ειδών που διαμορφώνει τις θεωρητικές προσδοχίες για τα αποτελέσματα που αναμένεται να προκύψουν σε κάθε πείραμα.

Το μοντέλο εφαρμόστηκε με κοινά γενικά χαρακτηριστικά αρχιτεκτονικής σε δύο μορφές αναπαράστασης.

8.1 Συνεχόμενες τιμές Offset, Quarterlength - One Hot Encoded Pitch

Στην πρώτη του μορφή το μοντέλο επιχείρησε την επίτευξη μεταφοράς μουσικού είδους με εκπαίδευσή πάνω σε δεδομένα μικτής αναπαράστασης. Συγκεκριμένα, όπως αναφέρθηκε στο 5ο κεφάλαιο, στο διάνυσμα αναπαράστασης νοτών (offset, quarterlength, pitch) χρησιμοποιήθηκαν συνεχόμενες τιμές για τις πρώτες δύο μεταβλητές, ενώ το pitch υποβλήθηκε σε κωδικοποίηση OHE.

Αν και εύχρηστο με γρήγορους ρυθμούς εκπαίδευσης, το μοντέλο δεν απέδωσε ικανοποιητικά στις μεταβλητές συνεχόμενων τιμών και αυτό, σε συνδυασμό με την δύσκολη αποκωδικοποίησή αποτελεσμάτων (οι προβλέψεις γίνονταν στο διάστημα $[0,1]$ και στη συνέχεια πολλαπλασιάζονταν με τη μέγιστη τιμή της ανάλογης μεταβλητής) ώθησε σε αναζήτηση εναλλακτικών.

8.2 One Hot Encoding των τριών μεταβλητών

Η κωδικοποίηση του διανύσματος με OHE ανάλογων διαστάσεων για κάθε μεταβλητή αποσκοπεί σε ακριβή πρόβλεψη τιμών χωρίς τα ζητήματα που προκύπτουν από συνεχόμενα διαστήματα αριθμών. Η προσέγγιση αυτή μειώνει την ταχύτητα εκπαίδευσης καθώς πολλαπλασιάζονται οι διαστάσεις που απαιτεί μια αναπαράστασή τιμής. Όπως προαναφέρθηκε, έχοντας θέσει όρια μεγίστου 600.0 και 50.0 στις μεταβλητές offset και quarterlength αντίστοιχα, μία τιμή τους αναπαρίσταται μέσω ενός πίνακα OHE διαστάσεων $\max*6 + 2$ όπου οι επιπλέον δύο όροι είναι το 0, που είναι εντός διαστήματος, και ο εκάστοτε όρος εκτός διαστήματος (έγινε επιλογή του $\max+0,25$) που χρησιμοποιείται για padding στα πλαίσια του teacher forcing. Εν τέλει ένα διάνυσμα μιας νότας έχει τις εξής διαστάσεις

(3.602, 302, 128)

Οι απαιτήσεις μνήμης της αναπαράστασης αυτής επιβάλλουν τη χρήση generator για τη σταδιακή φόρτωση των δεδομένων κατά την εκπαίδευση. Οι μεταβλητές batch_size και steps_per_epoch παρέμειναν σταθερές στα διάφορα στάδια των πειραμάτων στο 540 μετά από δοκιμές ως προς την ταχύτητα σύγκλισης και τη χρήση μνήμης.

Για την αξιολόγηση των OHE πειραμάτων χρησιμοποιήθηκε ο ταξινομητής κομματιών MIDI που αναπτύχθηκε στην εργασία 'Genre Recognition from Symbolic Music with CNNs' (Dervakos et al.)[29] πάνω στα evaluation sets, τα οποία έχουν μέγεθος περίπου 10% του αντίστοιχου train set κάθε είδους (85 κομμάτια κατά μέσο όρο). Ο ταξινομητής δίνει τη δυνατότητα κατηγοριοποίησης ενός MIDI κομματιού σε 12 συγκεκριμένα είδη προσφέροντας ένα ποσοστό ταυτοποίησης για κάθε κατηγορία. Η μέτρηση έγινε πριν και μετά την αλλαγή είδους σε κάθε κομμάτι και στη συνέχεια υπολογίστηκε ο μέσος όρος του κάθε set για το ενδιαφερόμενο είδος στην εκάστοτε φάση των πειραμάτων. Καθώς θεωρήθηκε σημαντική και η υποκειμενική συνεισφορά σε ένα τέτοιου είδους εγχείρημα ακολούθησε ανθρώπινη αξιολόγηση των πειραμάτων με διαφορετικές μεθόδους, σε όποια σημεία τα αποτελέσματα το επέτρεπαν.

Τα πειράματα με OHE μεταβλητές πραγματοποιήθηκαν σε τρία στάδια με σκοπό τη μείωση παραμέτρων που επηρεάζουν τον βαθμό δυσκολίας του εγχειρήματος.

8.2.1 Πρώτο στάδιο - Εκπαίδευση με ανεπεξέργαστα δεδομένα

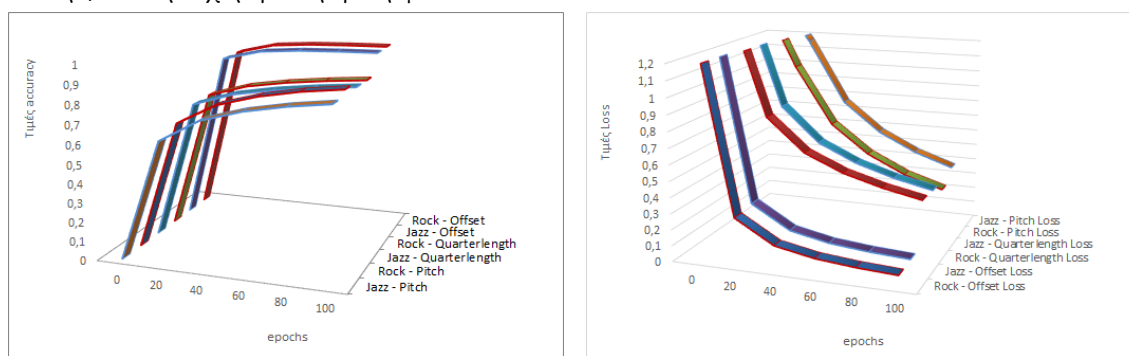
Στην πρώτη έκφανση των πειραμάτων τα δεδομένα φορτώθηκαν ως έχουν προς εκπαίδευση. Τα ζεύγη ειδών προς μεταφορά αποτέλεσαν τα Rock-Jazz και Rock-Electronic. Το ζεύγος Jazz-Electronic λόγω της πολυπλοκότητας και της μεγάλης διαφοράς των ειδών του εκπαιδευτήθηκε σε επόμενο στάδιο επεξεργασμένων, απλουστευμένων δεδομένων, όπου θεωρήθηκε ότι υπάρχουν αυξημένες πιθανότητες καλής απόδοσης.

Καθώς ο encoder των δύο train μοντέλων είναι κοινός ήταν σημαντικό το μέγεθος των δεδομένων κάθε είδους να είναι παρόμοιο σε κάθε ζεύγος, ώστε το γενικό μοντέλο να μην κλίνει περισσότερο σε ένα από τα δύο είδη. Με τον αριθμό να καθορίζεται από το είδος με τα λιγότερα κομμάτια, το Rock-Jazz πείραμα χρησιμοποίησε 688 και 673 κομμάτια που

απέφεραν 1,9 και 1,8 εκ. νότες αντίστοιχα, ενώ για το πείραμα Rock-Electronic υπήρχαν 999 και 925 κομμάτια διαθέσιμα για χρήση, αποφέροντας 3 και 2,8 εκ. νότες για τη δημιουργία των χρησιμοποιηθέντων διανυσμάτων.

Τα δύο train μοντέλα εκπαιδεύτηκαν εναλλασσόμενα ανά epoch όπως είθισται σε μοντέλα που έχουν κοινά δομικά στοιχεία για τη σταδιακή εκπαίδευσή τους στα δύο διαφορετικά set δεδομένων. Για το ζεύγος Rock-Jazz χρειάστηκαν 160 συνολικά epochs, 80 για κάθε ένα από τα δύο μοντέλα train, μέχρι να εμφανιστεί ικανοποιητική σύγκλιση και αποτελέσματα, ενώ το ζεύγος Rock-Electronic χρειάστηκε 200 epochs, 100 για την εκπαίδευσή κάθε είδους.

Παρατίθενται ενδεικτικά τα accuracy και loss graphs για αυτό το στάδιο πειραμάτων. Όπως ήταν αναμενόμενο το offset παρουσιάζει την ταχύτερη και υψηλότερη σύγκλιση από τις τρεις μεταβλητές λόγω της γραμμικά αυξανόμενης φύσης του, ενώ οι άλλες δυο συγκλίνουν πιο αργά και με χαμηλότερη ακρίβεια.



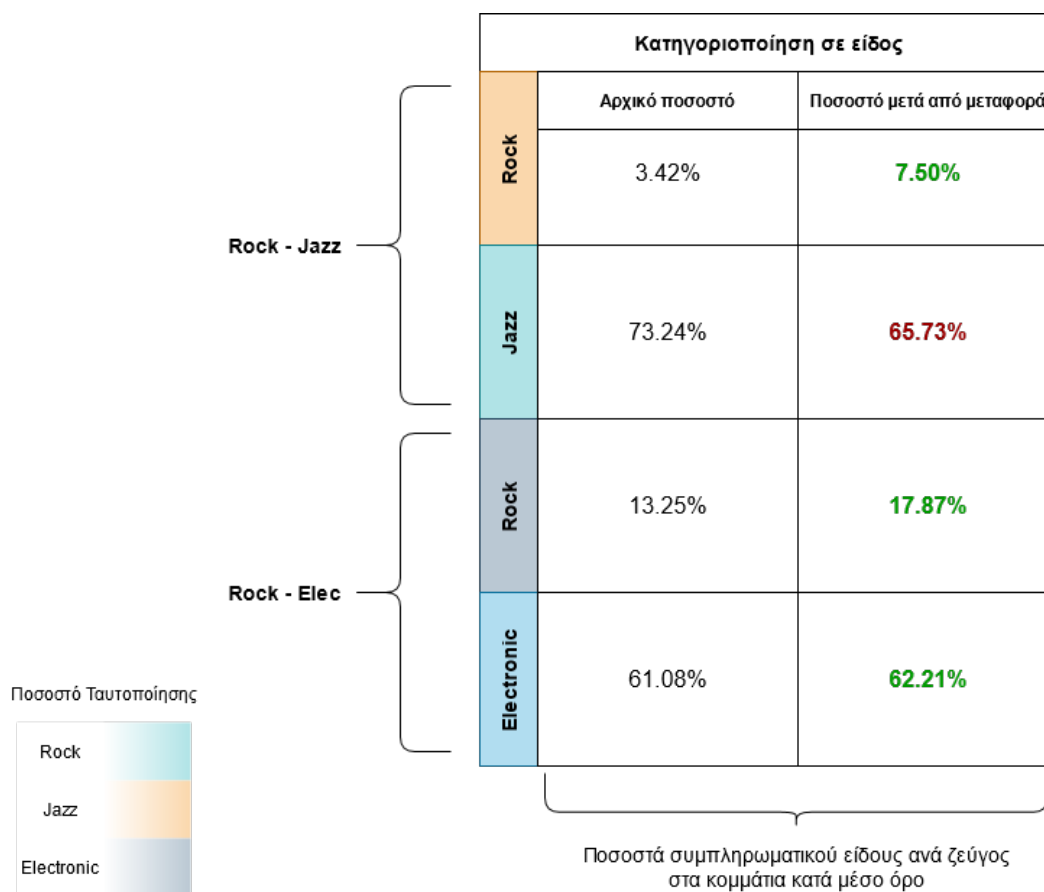
α': Accuracy

β': Loss

Σχήμα 8.1: Μετρικές για το μοντέλο Rock-Jazz του πρώτου σταδίου πειραμάτων

Ο ταξινομητής χρησιμοποιήθηκε για να μετρηθεί το ποσοστό ανίχνευσης του είδους μεταφοράς σε κομμάτια του αρχικού είδους του εκάστοτε ζεύγους. Για τις μετρήσεις χρησιμοποιήθηκαν κομμάτια που δεν αποτέλεσαν μέρος της εκπαίδευσής, πρώτα στην αρχική τους μορφή και έπειτα στο παραγόμενο από τη μεταφορά κομμάτι. Στο σχήμα 8.2 παρατίθενται τα αποτελέσματα των μετρήσεων.

Όπως φαίνεται τα κομμάτια της Rock παρουσίασαν την επιθυμητή αύξηση ταξινόμησης στις αντίστοιχες κατηγορίες μετά τη μεταφορά τους σε Jazz και Electronic, διπλασιάζοντας το ποσοστό κατηγοριοποίησης στην περίπτωση της Jazz. Αντίθετα, τα δύο πιο πολύπλοκα είδη δεν πέτυχαν ανάλογη μεταφορά στη Rock, με το ποσοστό της Jazz να μειώνεται σημαντικά κατά 10% και της Electronic να μένει οριακά στάσιμο παρουσιάζοντας μια μικρή αύξηση. Σύμφωνα με αυτές τις μετρήσεις, φαίνεται να υπάρχει μεγαλύτερη δυσκολία στη μεταφορά των ειδών με πιο συγκεκριμένα και ενδεχομένως πολύπλοκα χαρακτηριστικά σε ένα πιο απλό είδος από ό,τι το αντίστροφο. Αυτή η συμπεριφορά θα μπορούσε να θεωρηθεί σε ένα βαθμό αναμενόμενη με την παρούσα προσέγγιση όπου στο μοντέλο γίνεται παραποίηση των ρυθμικών στοιχείων και της συχνότητας μιας νότας αλλά δεν είναι εύκολη η συνειδητή αφαίρεσή της, μια κίνηση που μπορεί πιθανά να οδηγήσει στην απλούστευση ενός κομματιού. Παράγοντας αυτής της εικόνας φαίνεται να είναι και το υψηλό ποσοστό ανίχνευσης της Rock στις αρχικές μορφές των κομματιών Jazz και Electronic.



Σχήμα 8.2: Ποσοστά ταξινόμησης πρώτου σταδίου εκπαίδευσης σε ανεπεξέργαστα δεδομένα

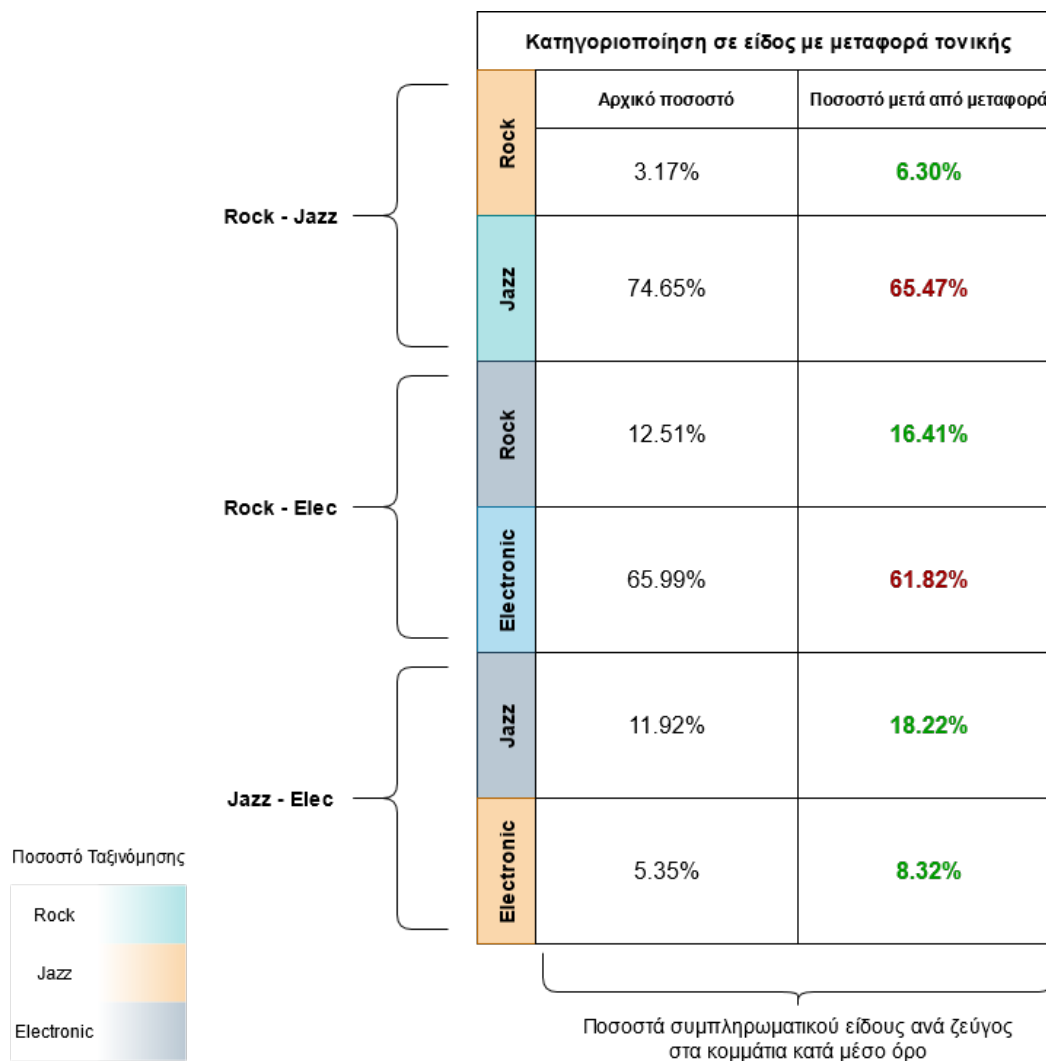
8.2.2 Δεύτερο στάδιο - Εκπαίδευση μετά από μεταφορά τονικότητας

Μια από τις μεγαλύτερες δυσκολίες στην εύρεση κοινών χαρακτηριστικών σύνθεσης ανάμεσα σε κομμάτια του ίδιου είδους, η οποία έχει θεωρητικά τη δυνατότητα να απλοποιηθεί, είναι η διαφορά τονικότητας μεταξύ κομματιών. Λόγω αυτής της παραμέτρου, κομμάτια που βρίσκονται σε κοινή κλίμακα και επομένως κινούνται σε μια κοινή αρμονική βάση, ενδεχομένως αντιμετωπίζονται από το μοντέλο ως πιο διαφορετικές περιπτώσεις από όσο είναι ορθό εάν έχουν διαφορετική τονικότητα. Με σκοπό την αποφυγή τέτοιας συμπεριφοράς από το μοντέλο, στο δεύτερο στάδιο των πειραμάτων εφαρμόστηκε αυτοματοποιημένη αναγωγή τονικού κέντρου σε ένα καθολικό (Nτο) για όλα τα κομμάτια των δεδομένων, με χρήση του αλγόριθμου Krumhansl-Schmuckler.

Αρχικά δεδομένα (ανά είδος)

Στο σύνολο των ζευγών ειδών προστέθηκε το Jazz-Electronic συμπληρώνοντας τους πιθανούς συνδυασμούς των Rock, Jazz, Electronic. Για το ζεύγος αυτό χρησιμοποιήθηκαν 673 και 675 κομμάτια των δύο ειδών αντίστοιχα, παρέχοντας 1,8 και 1,9 εκ. νότες προς εκπαίδευση.

Με τις ίδιες μεθόδους εναλλασσόμενων epoch, το νέο πείραμα Jazz-Electronic εκπαιδεύτηκε για 160 epochs συνολικά. Τα άλλα δύο ζεύγη εκπαιδεύτηκαν στους ίδιους αριθμούς με



Σχήμα 8.3: Ποσοστά ταξινόμησης δεύτερου σταδίου εκπαίδευσης σε δεδομένα κοινού τονικού κέντρου

το προηγούμενο στάδιο, μετά όμως από ενδείξεις overtraining για το πείραμα Rock-Electronic χρησιμοποιήθηκαν για αξιολόγηση τα μοντέλα των 160 epochs και σε αυτό το ζεύγος.

Σύμφωνα με τον ταξινομητή, παρά την προσπάθεια διευκόλυνσης του μοντέλου η απόδοσή του παραμένει στάσιμη σε σχέση με το πρώτο στάδιο των πειραμάτων (Σχήμα 8.3). Βασιζόμενοι σε αυτές τις μετρήσεις, συμπεραίνεται ότι οι πιθανές αλλαγές κλίμακας ή τονικότητας εντός ενός κομματιού, συμπεριφορές που δεν ήταν δυνατό να αποφευχθούν στο συγκεκριμένο dataset, εμπόδισαν την επίτευξη κοινής αρμονικής βάσης για όλα τα κομμάτια, παρουσιάζοντας στο μοντέλο ένα πεδίο προς εκπαίδευση παρόμοιο με αυτό των ανεπεξέργαστων δεδομένων. Η ελαφρώς χειρότερη απόδοση ενδεχομένως οφείλεται στη μικρή μείωση των νοτών προς εκπαίδευση κατά μερικές χιλιάδες μετά την μεταφορά τονικότητας στα κομμάτια, γεγονός που έχει να κάνει με τις μεθόδους του music21 που χρησιμοποιήθηκε για την μεταποίησή τους. Είναι επίσης πιθανό η μειωμένη αυτή απόδοση να οφείλεται σε πετυχημένη μεταφορά τονικότητας για ένα σημαντικό μέρος των dataset καθιστώντας δυσκολότερο για το μοντέλο τον χειρισμό

των λανθασμένων μεταφορών ως περιπτώσεων με μειωμένη πιθανότητα εμφάνισης. Οι μικρές διαφορές στα αρχικά ποσοστά σε σχέση με του πρώτου σταδίου οφείλονται στη χρήση των ίδιων κομματιών του evaluation set μετά από μεταφορά τονικότητας.

Στο πείραμα Jazz-Electronic παρουσιάζεται αύξηση των ενδείξεων για τις ανάλογες κατηγορίες μετά από την μεταφορά είδους και προς τις δύο κατευθύνσεις. Τα αποτελέσματα των άλλων μεταφορών υποδεικνύουν ότι μια ανάλογη απόδοση θα ήταν πιθανή και υπό τις συνθήκες του πρώτου σταδίου πειραμάτων.

Δεδομένα ανά στυλ

Στα πλαίσια της αναγωγής των αρμονικών κανόνων ενός συνόλου κομματιών σε ένα πιο κοινό επίπεδο με την μεταποίηση του τονικού κέντρου τους σε ένα καθολική, έγινε μία προσπάθεια διεύρυνσης του εγχειρήματος στη μεταφορά στυλ πέρα από το είδος. Πιο συγκεκριμένο στην υπόθεση, αποτελεί προσέγγιση που θα είχε αρκετή διευκόλυνση στα πλαίσια του δεύτερου σταδίου πειραμάτων.

Χρησιμοποιώντας κομμάτια του dataset τα οποία είχαν ταυτοποιηθεί και με labels ύφους εκτός από είδους, δημιουργήθηκε το ζεύγος μεταφοράς στυλ Happy-Sad για την μεταποίηση κομματιών που έχουν χαρακτηριστεί εύθυμα ή στενάχωρα. Οι δύο κατηγορίες κομματιών αποτελούνταν από περίπου 1000 κομμάτια η κάθε μία και εκπαιδεύτηκαν με τον ίδιο τρόπο που έγινε εκπαίδευση σε ένα ζεύγος ειδών, για 200 epochs συνολικά.

Καθώς η αναγωγή τονικότητας σε κοινή δεν ήταν επιτυχής σε ικανοποιητικό βαθμό όπως φάνηκε από την αντίστοιχη απόδοση στην μεταφορά είδους, δεν σημειώθηκε το αναμενόμενο αποτέλεσμα μεταποίησης ενός εύθυμου κομματιού σε στενάχωρο ή του αντιστρόφου. Χωρίς την κοινή αρμονική βάση το εγχείρημα αυτό δυσκολεύει σημαντικά για αρχεία MIDI, καθώς πέρα από τη χρήση συγκεκριμένων κλιμάκων η κατηγοριοποίηση κομματιών σε αυτά τα στυλ προκύπτει κυρίως από την χροιά που τα χαρακτηρίζει, μεταβλητή η οποία δεν περιλαμβάνεται σε συμβολικές αναπαραστάσεις.

8.2.3 Τρίτο στάδιο - Εκπαίδευση μετά από απαλοιφή του Pitch


Τέλος, απλοποιώντας το διάνυσμα αναπαράστασης σε πληροφορίες σχετικές μόνο με το χρονικό σημείο που προκύπτει μια νότα και τη διάρκειά της επιχειρήθηκε να αναδειχθούν οι ρυθμικές διαφορές των τριών ειδών οι οποίες ίσως είναι και οι πιο προφανείς.

Με την απλούστερη αυτή αναπαράστασή χρειάστηκαν μόνο 40 epochs συνολικά για να συγκλίνουν οι εκπαιδεύσεις πάνω σε ίδιο αριθμό από instances με τα αντίστοιχα ζεύγη ειδών στα προηγούμενα στάδια. Για τις προβλέψεις, τα παραγόμενα δισδιάστατα διανύσματα ενώθηκαν με τα pitch στα οποία αντιστοιχούσαν για να παρατηρηθούν οι διαφορές με τα αρχικά κομμάτια.

Η απαλοιφή του Pitch δεν απέδωσε αποτελέσματα με κάποιο ενδιαφέρον με τα κομμάτια να αξιολογούνται από τον ταξινομητή με περίπου ίδιες βαθμολογίες πριν και μετά την μεταφορά. Μπορεί να ληφθεί το συμπέρασμα ότι στη συγκεκριμένη κατάσταση μοντέλου η απουσία του

		Κατηγοριοποίηση σε είδος με απαλοιφή pitch	
		Αρχικό ποσοστό	Ποσοστό μετά από μεταφορά
Rock - Jazz	Rock	3.42%	3.83%
	Jazz	73.24%	67.95%
Rock - Elec	Rock	13.25%	12.35%
	Electronic	61.08%	62.75%
Jazz - Elec	Jazz	12.71%	14.04%
	Electronic	3.72%	5.57%

Ποσοστά συμπληρωματικού είδους στα κομμάτια κατά μέσο όρο



Σχήμα 8.4: Ποσοστά ταξινόμησης τρίτου σταδίου εκπαίδευσης σε δεδομένα με απαλοιφή του pitch

pitch αποτρέπει την διαχώριση μεταξύ ειδών οδηγώντας σε απλή λειτουργία του ως autoencoder ανεξαρτήτως του decoder που χρησιμοποιείται.

Δεν είναι απίθανη και η εκδοχή του overtrain ακόμα και στον ιδιαίτερα χαμηλό αριθμό epochs, κάτι που δεν υπήρχε χρόνος να εξακριβωθεί με την επανάληψή του πειράματος στα πλαίσια αυτής της εργασίας.

8.3 Επιπλέον αξιολόγηση

Λόγω της υποκειμενικής φύσης που χαρακτηρίζει εξ ορισμού το γενικότερο εγχείρημα θεωρήθηκε κομβική η ανθρώπινη αξιολόγηση των αποτελεσμάτων πέρα από τη χρήση αυτοματοποιημένων μεθόδων. Για το λόγο αυτό έγιναν πολλαπλές εφαρμογές των διαφόρων μοντέλων σε συγκεκριμένα κομμάτια, από train και evaluation sets, που επιλέχθηκαν χειροκίνητα με γνώμονα την εύστοχη εκπροσώπηση ενός είδους και τις πρακτικές δυνατότητες

μεταφοράς τους σε ένα άλλο. Η διάκριση αυτών των στοιχείων έγινε ηχητικά, αναπαράγοντας τα κομμάτια με όσο το δυνατόν πιο ουδέτερα στοιχεία χροιάς, η οποία είναι πλέον παρούσα, έτσι ώστε οι διακρίσεις των τριών ειδών να είναι όσο γίνεται πιο αντικειμενικές.

Σε ένα μεγάλο βαθμό η τακτική αυτή επιβεβαιώνει την γενική εικόνα που παρουσιάζει ο ταξινομητής, με κάποιες ουσιαστικές ωστόσο διαφορές. Η βασικότερη εξ αυτών αφορά το δεύτερο στάδιο πειραμάτων όπου έγινε μεταφορά τονικότητας για τα κομμάτια του dataset. Αν και όπως υποδεικνύεται από τις μετρήσεις, τα αποτελέσματα δεν παρουσίασαν σημαντικά μεγαλύτερη αρμονική συνοχή από ό,τι στο πρώτο στάδιο, διακρίθηκε, ιδίως στη μεταφορά κομματιών Rock σε Electronic, μία τάση εισαγωγής νέων μουσικών φράσεων με μελωδική συνέπεια στο ανάλογο είδος, και εύηχης μεταποίησης των αρμονικών βάσεων του αρχικού κομματιού. Στο πρώτο στάδιο πειραμάτων, τα αντίστοιχα σημεία παρουσίασαν κυρίως ρυθμική μεταποίηση ενώ οι αλλαγές στο pitch δεν σημειώνονταν σε βαθμό σχηματισμού νέας συνεκτικής μουσικής φράσης. Κατά κύριο λόγο, οι ρυθμικές μεταφορές ήταν εμφανείς από το αρχικό στάδιο πειραμάτων, με εύστοχες αλλαγές από τη Rock στα άλλα δύο είδη. Αλλαγές είδους σε κομμάτια του train dataset απέδωσαν τα πιο ενδιαφέροντα από μουσικής απόψεως αποτελέσματα, παρουσιάζοντας ανά σημεία εύηχες μουσικές φράσεις με τα προσδοκώμενα στοιχεία.

Τα προγράμματα που χρησιμοποιήθηκαν για την επεξεργασία δεδομένων εισόδου και εξόδου, αλλά και αυτά τα οποία υλοποίησαν τις εκπαιδεύσεις γράφτηκαν σε γλώσσα προγραμματισμού Python. Για τον χειρισμό των δεδομένων σε διαφορετικές μορφές χρησιμοποιήθηκε εκτενώς το music21[30] το οποίο αποτελεί ένα Python-based toolkit που αναπτύχθηκε από το μουσικό τμήμα του MIT για μουσικολογία με βοήθεια υπολογιστή.

Ο σχεδιασμός και η εκπαίδευση των μοντέλων πραγματοποιήθηκαν με τη χρήση της βιβλιοθήκης Keras[31] η οποία παρέχει ένα περιβάλλον σε Python για την ανάπτυξη νευρωνικών δικτύων. Το Keras λειτουργεί ως διεπαφή για την γενικότερη βιβλιοθήκη Tensorflow[32] η οποία έχει αναπτυχθεί από την Google για μηχανική μάθηση. Πέρα από τον εφοδιασμό με συναρτήσεις ευρέως χρησιμοποιούμενες για την εκπαίδευση μοντέλων και την μέτρηση απόδοσής τους, οι βιβλιοθήκες παρείχαν τη δυνατότητα χρήσης GPU για την αύξηση των ταχυτήτων εκπαίδευσης σε σχέση με το αργότερων ρυθμών CPU. Η διεπαφή CUDA[33] χρησιμοποιήθηκε μέσω Tensorflow για την αξιοποίηση των διαθέσιμων NVIDIA GPU.

10.1 Σφάλματα

Οι προσεγγίσεις που εφαρμόστηκαν εντός της παρούσας εργασίας εξαρτώνται άμεσα από κάποιες θεωρητικές υποθέσεις σχετικά με τη λειτουργικότητα και την ορθή οργάνωση των στοιχείων που χρησιμοποιούνται. Συγκεκριμένα, κατά την οργάνωση ενός μεγάλου dataset όπως το χρησιμοποιηθέν Lakh είναι αδύνατον να αποφευχθούν οι εσφαλμένες κατηγοριοποιήσεις κομματιών λόγω τεχνικών λαθών. Κάτι τέτοιο δυσκολεύει την ικανοποιητική σκιαγράφηση ενός είδους και αποπροσανατολίζει τη σύγκρισή του με άλλα είδη λόγω της παρουσίας κοινών κομματιών σε set προς εκπαίδευση διαφορετικών ειδών. Το συγκεκριμένο dataset παρέχει σχετική διευκρίνιση για λάθη στα label που παρέχει, παραμένει όμως η πιο χρηστική επιλογή για τον μαζικό χειρισμό MIDI αρχείων. Επίσης, ρόλο για το συγκεκριμένο εγχείρημα παίζει και η υποκειμενικότητα που υπήρξε στον χαρακτηρισμό των κομματιών από τις πηγές που χρησιμοποιήθηκαν για εξαγωγή πληροφοριών.

Εξίσου μεγάλος παράγοντας ήταν το εύρος των τριών ειδών. Η Rock, Jazz και Electronic αποτελούν γενικούς όρους που καλύπτουν ένα μεγάλο φάσμα υποειδών τα οποία σε ορισμένες περιπτώσεις διαφέρουν σημαντικά στην προσέγγιση σύνθεσης. Το υποείδος που υπερισχύει σε κάθε dataset μπορεί να επηρεάζει τις θεωρητικές προσδοκίες συμπεριφοράς του μοντέλου, όπως ενδεχομένως συνέβη στην περίπτωση της Jazz, όπου διαπιστώθηκε ότι μεγάλος αριθμός των κομματιών αναφερόταν σε πρώιμα παρακλάδια του είδους με έντονα στοιχεία blues, τα οποία δεν εμφανίζουν την ανάλογη πολυπλοκότητα που συνδέεται με το είδος σήμερα. Ο μειωμένος αριθμός διαθέσιμων κομματιών για πιο ειδικά υποείδη δεν επέτρεψε μια συγκεκριμενοποίηση της προσέγγισης ως προς το είδος και γενίκευσε το πεδίο αναμενόμενων αποτελεσμάτων εις βάρος του μοντέλου. Εξαίρεση αποτελεί το set κομματιών Electronic που προέκυψε από συνδυασμό υποειδών ηλεκτρονικής μουσικής με κοντινά χαρακτηριστικά.

Η μεταφορά τονικότητας που επιχειρήθηκε κατά το δεύτερο σκέλος των πειραμάτων φάνηκε εκ του αποτελέσματος ότι επίσης δεν εφαρμόστηκε υπό ιδανικές συνθήκες. Η εξασφάλιση σταθερής κλίμακας και τονικού κέντρου ανά κομμάτι ενδεχομένως θα εξασφάλιζε επιτυχή εντοπισμό και αναγωγή τονικότητας και κατά συνέπεια βελτιωμένα αποτελέσματα σε σχέση με τα άλλα στάδια πειραμάτων, επιβεβαιώνοντας πρακτικά την θεωρητική λειτουργικότητα της

ιδέας.

10.2 Συμπεράσματα

Η διεξαγωγή των πειραμάτων για τη μεταφορά είδους σε αρχεία MIDI απέφερε έναν αριθμό επιθυμητών αποτελεσμάτων τα οποία θα μπορούσαν να έχουν σημειώσει σταδιακή βελτίωση με τις απλοποιήσεις αναπαραστάσεων δεδομένων που εφαρμόστηκαν. Τα πειράματα κατέδειξαν τα σημεία που χρειάζονται επαναπροσδιορισμό προκειμένου να ανταποκριθεί πιο δυναμικά το μοντέλο που σχεδιάστηκε.

Το δεύτερο στάδιο πειραμάτων έμεινε στάσιμο σε απόδοση λόγω των τεχνικών περιορισμών στη μεταφορά τονικότητας που προαναφέρθηκαν, ενώ στο τρίτο η περαιτέρω απλοποίηση των δεδομένων ενδεχομένως απέκοψε τις απαραίτητες πληροφορίες που διακρίνουν τα είδη για το μοντέλο, οδηγώντας σε αμελητέα μεταποίηση των κομματιών. Οι αποδόσεις αυτές μπορούν να κατευθύνουν σχετικές έρευνες που θα επιχειρήσουν παρόμοιο εγχείρημα με στοχευμένες προσεγγίσεις για την ολοκλήρωση και την διεύρυνση των προτεινόμενων πρακτικών.

10.3 Μελλοντική Έρευνα

Η ανολοκλήρωτη μορφή των δύο τελευταίων σταδίων πειραμάτων αφήνει ανοιχτές δυνατότητες υψηλής βελτίωσης σε πεδία που κρίνονται πρόσφορα για το εν λόγω εγχείρημα. Η πετυχημένη αναγωγή τονικής είτε με χρήση καταλληλότερων δεδομένων με σταθερά αρμονικά χαρακτηριστικά, ή με εύρεση αποδοτικότερων τρόπων αρμονικής ανάλυσής τους, μπορεί να θεωρηθεί απαραίτητη για την αποδοτικότερη μεταφορά είδους και στυλ. Ενδιαφέροντα αποτελέσματα ενδέχεται να παρουσιάσουν και παραλλαγμένες ρυθμικές προσεγγίσεις, καθώς η αδυναμία απόδοσης της υπάρχουσας δεν αναιρεί τη σημασία των ρυθμικών στοιχείων στο διαχωρισμό ειδών όπως έχει αναφερθεί επανειλημμένα σε άλλες έρευνες.

Κατά την οργάνωση της προσέγγισης που επιλέχθηκε υπήρξαν στοιχεία που παρακάμφθηκαν καθώς η συμπερίληψή τους θα ήταν πολλαπλασιαστική στον βαθμό δυσκολίας. Κάποια από αυτά, όπως το tempo, θα μπορούσαν να παρέχουν ωφέλιμες πληροφορίες κατά την εκμάθηση του μοντέλου πάνω σε διαφορετικά είδη, μέσω μιας αναπαράστασης που τα περιελάμβανε. Ιδιαίτερα πιο πολύπλοκη αλλά αδιαμφισβήτητα κομβική προέκταση θα ήταν κωδικοποίηση των MIDI tracks κρουστών σε μια μορφή εύχρηστη καθώς τα περιεχόμενά τους πολλές φορές αρκούν για την αναγνώριση ενός είδους. Μια τέτοια προσέγγιση που μπαίνει στη διαδικασία διαχωρισμού ρυθμικών και μελωδικών track αντί για τη συμπίεση των αποκλειστικά μελωδικών σε ένα καθολικό, θα μπορούσε να επεκταθεί περαιτέρω στη διατήρηση δομής του GM 1, εξάγοντας κάθε track ενός κομματιού ξεχωριστά και επισυνάπτοντας την πληροφορία του οργάνου στο οποίο αναφέρεται. Συνδυάζοντας αυτά τα στοιχεία προκύπτει μια συμβολική αναπαράσταση κομματιού η οποία περιέχει αναφορές και στη θεωρητική χροιά του. Αυτό το εγχείρημα θα παρείχε σημαντικές διευκολύνσεις και υψηλής πιστότητας αποτελέσματα στην μεταφορά ηχητικών αρχείων από είδος σε είδος.

Η δημιουργία ενός εργαλείου μεταποίησης μεταξύ εύχρηστων αρχείων συμβολικής αναπαράστασης μουσικής είναι εφικτή και η συνέχιση των προσπαθειών για την ολοκλήρωσή του είναι σημαντική για τη σύγχρονη διαδικασία μουσικής σύνθεσης.

Βιβλιογραφία

- [1] Michael Mateas. Interactive Drama, Art and Artificial Intelligence (2002) <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.66.807>
- [2] Wikipedia. Music and artificial intelligence https://en.wikipedia.org/wiki/Music_and_artificial_intelligence, visited June 2021
- [3] Jeremy Freeman <https://medium.com/@jeremy.freeman.53491/artificial-intelligence-and-music-what-the-future-holds-79005bba7e7d>, visited June 2021
- [4] Beard D., Gloag K. Musicology - The Key Concepts (1987) https://www.academia.edu/4362932/Beard_D_Gloag_K_Musicology_The_Key_Concepts
- [5] Cory McKay. Automatic Genre Classification of MIDI Recordings (2004) <http://www.music.mcgill.ca/~cmckay/papers/musictech/MA.Thesis.pdf>
- [6] David Miles Huber. The MIDI manual (2007)
- [7] Hao-Wen Dong, Wen-Yi Hsiao, and Yi-Hsuan Yang, “Pypianoroll: Open Source Python Package for Handling Multitrack Pianorolls,” in Late-Breaking Demos of the 19th International Society for Music Information Retrieval Conference (ISMIR) (2018) <https://salu133445.github.io/pypianoroll/pdf/pypianoroll-ismir2018-lbd-paper.pdf>
- [8] Stephen Mcadams. Musical Timbre Perception (2013) https://www.researchgate.net/publication/279946150_Musical_Timbre_Perception
- [9] Julia C. Hailstone et al. It’s not what you play, it’s how you play it: Timbre affects perception of emotion in music (2009) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2683716/>
- [10] Hajda, J. M., Kendall, R. A., Carterette, E. C., Harshberger, M. L. Methodological issues in timbre research (1997) <https://psycnet.apa.org/record/1997-09141-012>
- [11] Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang, “MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment,” in Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI), 2018
- [12] Colin Raffel, “Learning-Based Methods for Comparing Sequences, with Applica-

tions to Audio-to-MIDI Alignment and Matching,” PhD Thesis, 2016

[13] Jason Brownlee. Text Generation With LSTM Recurrent Neural Networks in Python with Keras <https://machinelearningmastery.com/text-generation-lstm-recurrent-neural-networks-python-keras/>, visited June 2021

[14] Krumhansl, C. L. Cognitive foundations of musical pitch (1990) <https://psycnet.apa.org/record/2004-14155-000>

[15] Gouyon, Fabien Dixon, Simon Pampalk, Elias Widmer, Gerhard. Evaluating rhythmic descriptors for musical genre classification (2004) https://www.researchgate.net/publication/228489438_Evaluating_rhythmic_descriptors_for_musical_genre_classification

[16] Esparza, Tlacacl Bello, Juan Humphrey, Eric. From Genre Classification to Rhythm Similarity: Computational and Musicological Insights. Journal of New Music Research. 44. 39-57. 10.1080/09298215.2014.929706. (2014) https://www.researchgate.net/publication/273296604_From_Genre_Classification_to_Rhythm_Similarity_Computational_and_Musicological_Insights [17] Jesse Engel et al. Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders (2017) <https://arxiv.org/abs/1704.01279>

[18] Noam Mor et al. A Universal Music Translation Network (2018) <https://arxiv.org/abs/1805.07848>

[19] Ilya Sutskever et al. Sequence to Sequence Learning with Neural Networks (2014) <https://arxiv.org/abs/1409.3215>

[20] Ronald J. Williams et al. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks (1989) <https://arxiv.org/abs/1409.3215>

[21] Nair, Vinod Hinton, Geoffrey. Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair. Proceedings of ICML. 27. 807-814. (2010) https://www.researchgate.net/publication/221345737_Rectified_Linear_Units_Improve_Restricted_Boltzmann_Machines_Vinod_Nair

[22] Jun Han, Claudio Moraga. The influence of the sigmoid function parameters on the speed of backpropagation learning (2005) <https://psycnet.apa.org/record/2004-14155-000>

[23] Jun Han, Claudio Moraga. The influence of the sigmoid function parameters on the speed of backpropagation learning (2005) <https://psycnet.apa.org/record/2004-14155-000>

[24] Jason Brownlee. How to Choose an Activation Function for Deep Learning <https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/>, visited June 2021

[25] Jason Brownlee. How to Choose Loss Functions When Training Deep Learning Neural Networks <https://machinelearningmastery.com/how-to-choose-loss-functions-when-training-deep-learning-neural-networks/>, visited June 2021

[26] Github Keras team. <https://github.com/keras-team/keras/issues/6444>, visited June 2021

[27] Diederik P. Kingma et al. Adam: A Method for Stochastic Optimization (2010) <https://arxiv.org/pdf/1412.6980.pdf>

[28] Aleksey Bilogur. Keras optimizers (2018) <https://www.kaggle.com/residentmario/keras-optimizers>, visited June 2021

[29] Edmund Dervakos, Natalia Kotsani, Giorgos Stamou. Genre Recognition from Symbolic Music with CNNs (2021) https://link.springer.com/chapter/10.1007/978-3-030-72914-1_7

[30] Michael Scott Cuthbert, Christopher Ariza. music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data (2010) <https://arxiv.org/pdf/1412.6980.pdf>

[31] Keras: The Python Deep Learning Library <https://keras.io/>

[32] Tensorflow, An open source machine learning framework for everyone <https://www.tensorflow.org/>

[33] CUDA compute platform, NVIDIA <https://developer.nvidia.com/about-cuda>

