Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Πολυτεχνική Σχολή

Τμήμα Ηλεκτρολόγων Μηχανικών
& Μηχανικών Υπολογιστών

ΑΝΑΓΝΩΡΙΣΗ ΜΟΥΣΙΚΟΥ ΕΙΔΟΥΣ

Μόσχος Σωτήριος (9030)

Καλαντζής Γεώργιος (8818)

Παπακώστας Γεράσιμος (8890)

Περίληψη

Η παρούσα εργασία πραγματοποιήθηκε στο πλαίσιο του μαθήματος Τεχνολογία Ήχου και Εικόνας, του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης. Η ανάπτυξη του διαδικτύου έχει οδηγήσει τη μουσική βιομηχανία σε μια μετάβαση από τα φυσικά μέσα σε διαδικτυακά προϊόντα και υπηρεσίες. Άμεση συνέπεια της παραπάνω μετάβασης είναι η διαδικτυακή αποθήκευση μουσικών συλλογών, οι οποίες εμπλουτίζονται διαρκώς με χιλιάδες νέα μουσικά κομμάτια. Επομένως, έχει δημιουργηθεί η ανάγκη για μουσικές τεχνολογίες, οι οποίες θα επιτρέπουν στους χρήστες να έχουν πρόσβαση σε αυτές τις εκτενείς συλλογές με αποτελεσματικό και αποδοτικό τρόπο.

Η συγκεκριμένη εργασία αποτελεί μια προσπάθεια για την αυτόματη ταξινόμηση των μουσικών κομματιών ανάλογα με το είδος στο οποίο ανήκουν. Πιο συγκεκριμένα, χρησιμοποιείται ως πηγή το μουσικό σήμα, από το οποίο εξάγουμε συγκεκριμένα χαρακτηριστικά. Κατόπιν, πραγματοποιείται μια επιλογή ορισμένων χαρακτηριστικών από τα εξαγόμενα με τη χρήση πολλαπλών τεχνικών, ώστε να χρησιμοποιηθούν ως είσοδος στα μοντέλα μηχανικής μάθησης με επίβλεψη που αναπτύχθηκαν, με σκοπό την καλύτερη δυνατή ταξινόμηση των μουσικών κομματιών.

Η ιδιαιτερότητα της είναι ότι πέρα από τις κλασσικές τεχνικές επιλογής χαρακτηριστικών και μεθόδων ταξινόμησης που υλοποιήθηκαν, παρουσιάζεται και μια εναλλακτική πρόταση χρονικής συνάθροισης χαρακτηριστικών (temporal feature integration) και συγχώνευσης των αποφάσεων των ταξινομητών (fusion of decisions), με σκοπό την επίτευξη βέλτιστης ταξινόμησης των μουσικών κομματιών.

Για την παρούσα εργασία, χρησιμοποιήθηκε το GTZAN σύνολο δεδομένων, το οποίο αποτελείται από 1000 μουσικά κομμάτια, που ταξινομήθηκαν σε 10 διαφορετικά μουσικά είδη: Μπλουζ, κλασσική μουσική, κάουντρι, ντίσκο, χιπ-χοπ, τζαζ, μέταλ, ποπ, ρέγκε και ροκ.

Μόσχος Σωτήριος, moschoss@ece.auth.gr

Καλαντζής Γεώργιος, gkalantz@ece.auth.gr

Παπακώστας Γεράσιμος, gpapakos@ece.auth.gr

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών,

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Ελλάδα

Νοέμβριος 2020

Επίβλεψη

Δημούλας Α. Χαράλαμπος, Αναπληρωτής Καθηγητής

Θωίδης Ιορδάνης, Υποψήφιος Διδάκτορας

Ευρετήριο σχημάτων

Σχήμα 1: Αποτελέσματα προσέγγισης με χρήση παραδοσιακών μεθόδων μηχανικής μάθησης 13

Σχήμα 2: Αποτελέσματα προσέγγισης με τεχνική χρονικής συνάθροισης χαρακτηριστικών..... 14

Περιεχόμενα

1 **Εισαγωγή** 4

1.1 Κίνητρο και σημασία του θέματος 4

1.2 Δομή εργασίας 5

2 **Προσέγγιση του προβλήματος** 5

2.1 Παραδοσιακές τεχνικές μηχανικής μάθησης 5

2.1.1 Σύνολο δεδομένων 6

2.1.2 Εξαγωγή χαρακτηριστικών 6

2.1.3 Προεπξεργασία 8

2.1.4 Επιλογή/Μετασχηματισμός χαρακτηριστικών 8

2.1.5 Ταξινόμηση 9

2.2 Τεχνική χρονικής συνάθροισης χαρακτηριστικών 10

3 **Βιβλιογραφική επισκόπηση** 11

4 **Μεθοδολογίες** 12

4.1 Μεθοδολογία πρώτης προσέγγισης 12

4.2 Μεθοδολογία δεύτερης προσέγγισης 13

5 **Αποτελέσματα** 14

5.1 Αποτελέσματα προσέγγισης με παραδοσιακές τεχνικές μηχανικής μάθησης 14

5.2 Αποτελέσματα προσέγγισης με τεχνική χρονικής συνάθροισης χαρακτηριστικών 15

6 **Συμπεράσματα και μελλοντικές προεκτάσεις** 16

6.1 Συμπεράσματα 16

6.2 Μελλοντικές προεκτάσεις 16

Κεφάλαιο 1

1. Εισαγωγή

1.1 Κίνητρο και σημασία του θέματος

Το αντικείμενο της μουσικής αποτελεί ένα παγκόσμιο φαινόμενο, που μελετάται, δημιουργείται και απολαμβάνεται από ένα ευρύ και ποικίλο κοινό. Στην εποχή της ψηφιακής πληροφορίας, η πλειονότητα των μουσικών κομματιών έχει γίνει διαθέσιμη σε κάθε χρήστη.

Ωστόσο, η αφθονία της πληροφορίας είναι τόσο μεγάλη και διαφορετική, κάτι το οποίο την καθιστά δύσκολα διαχειρίσιμη, πράγμα που οδηγεί στην ανάγκη ανάπτυξης αυτομάτων τεχνικών για την ταξινόμησή της σε συλλογές, με μια από αυτές να αποτελεί το μουσικό είδος στο οποίο ανήκουν τα μουσικά κομμάτια.

Το συγκεκριμένο πρόβλημα αποτελεί αντικείμενο μελέτης εδώ και αρκετά χρόνια και για την επίλυσή του έχει αναπτυχθεί μεγάλος αριθμός τεχνικών, οι οποίες μπορούν να κατηγοριοποιηθούν ως εξής:

- Εξαγωγή μουσικών χαρακτηριστικών (audio features), τα οποία βασίζονται στην επεξεργασία του μουσικού σήματος ενός κομματιού και χρησιμοποίηση αυτών για την εκπαίδευση ενός μοντέλου κλασσικής μηχανικής μάθησης.
- Εισαγωγή ολόκληρου του μουσικού σήματος σε ένα μοντέλο βαθιάς μάθησης (deep learning).

Η συγκεκριμένη εργασία βασίζεται στο σκεπτικό της πρώτης κατηγορίας τεχνικών που αναφέρθηκαν παραπάνω και έχει ως σκοπό την εφαρμογή μια νέας τεχνικής που θα προσδώσει αποτελεσματικότερη επίλυση του συγκεκριμένου προβλήματος.

1.2 Δομή Εργασίας

Το παρόν κεφάλαιο αποτελεί μια εισαγωγή στη σημασία του θέματος, το στόχο της εργασίας και παρουσιάζει την δομής της. Πιο αναλυτική, ακολουθεί την παρακάτω δομή:

Στο **δεύτερο** κεφάλαιο, γίνεται μια παρουσίαση της προσέγγισης του προβλήματος και των μεθόδων που αναπτύχθηκαν για την επίλυσή του.

Στο **τρίτο** κεφάλαιο, πραγματοποιείται μια επισκόπηση των δημοσιεύσεων και πηγών που μελετήθηκαν.

Στο **τέταρτο** κεφάλαιο, παρουσιάζονται οι μεθοδολογίες, και συγκεκριμένα οι δύο διαφορετικές προσεγγίσεις που ακολουθήθηκαν.

Στο **πέμπτο** κεφάλαιο, γίνεται μια σύνοψη των αποτελεσμάτων των διαφόρων μοντέλων.

Στο **έκτο** κεφάλαιο, πραγματοποιείται μια παρουσίαση μελλοντικών επεκτάσεων και βελτιώσεων που μπορούν να βελτιστοποιήσουν ακόμη περισσότερο την παρούσα προσέγγιση.

Κεφάλαιο 2

2. Προσέγγιση του προβλήματος

2.1 Παραδοσιακές τεχνικές μηχανικής μάθησης

Η αρχική προσέγγιση επίλυσης του προβλήματος ακολούθησε την παρακάτω μεθοδολογία:

- Εξαγωγή χαρακτηριστικών από το πεδίο χρόνου και από το πεδίο της συχνότητας του μουσικού σήματος.
- Προεπεξεργασία και ανάλυση χαρακτηριστικών.
- Επιλογή/Μετασχηματισμός χαρακτηριστικών βάσει συγκεκριμένων κριτήριων.
- Εισαγωγή χαρακτηριστικών σε διάφορα μοντέλα ταξινόμησης μηχανικής μάθησης, με αποτέλεσμα την αναγνώριση του επιθυμητού είδους.

2.1.1 Σύνολο δεδομένων

Για τη συγκεκριμένη εργασία χρησιμοποιήθηκε το GTZAN σύνολο δεδομένων, που αποτελεί ένα ευρέως χρησιμοποιούμενο συνόλο για το συγκεκριμένο πρόβλημα. Τα αρχεία ήχου συλλέχθηκαν από το 2000 έως το 2001 από μια πληθώρα μουσικών πηγών.

Αποτελείται από 1000 μουσικά κομμάτια διάρκειας 30 δευτερολέπτων τα οποία κατηγοριοποιούνται σε 10 μουσικά είδη. Διαθέσιμος υπερσύνδεσμος:

http://marsyas.info/downloads/datasets.html

2.1.2 Εξαγωγή χαρακτηριστικών

Η εξαγωγή χαρακτηριστικών προκύπτει από την ανάλυση του ηχητικού σήματος στο πεδίο του χρόνου και της συχνότητας. Αρχικά, στο πεδίο του χρόνου πραγματοποιήθηκε εξαγωγή των παρακάτω χαρακτηριστικών:

- Μέσος όρος του πλάτους του σήματος
- Τυπική απόκλιση του πλάτους του σήματος
- Στρέβλωση του πλάτους του σήματος
- Κυρτότητα του πλάτους του σήματος
- Ρυθμός μετάβασης από το σημείο(0,0) του διακριτού σήματος x(n) (Gouyon et al., 2000)

  $$Z_n = \sum_m |sgn[x(m)] - sgn[x(m-1)]|\, w(n-m) \quad (1)$$

  όπου,

  $$sgn[x(m)] = \begin{cases} -1, & x(n) < 0 \\ 1, & x(n) \geq 0 \end{cases}$$

  και w(n) ορθογώνιο παράθυρο συνάρτησης,

  $$w(n) = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & \alpha\lambda\lambda o\acute{\upsilon} \end{cases}$$

  όπου, N το μήκος του παραθύρου.

- Τετραγωνική ρίζα μέσης ενέργειας του σήματος

  $$E_n = \sqrt{\sum_m [x(m)w(n-m)]^2} \quad (2)$$

  όπου, x(n), w(n) οι ίδιες μεταβλητές με παραπάνω.

- Τέμπο, με αυτόν τον όρο αναφερόμαστε στην περιοδική αναπαραγωγή του ήχου σε τακτά χρονικά διαστήματα. (Grosche et al. ,2010)

Στο πεδίο της συχνότητας για να εξαχθούν χαρακτηριστικά εκμεταλλευόμαστε τον μετασχηματισμό Φουριέ Βραχέου Χρόνου (Short-Time Fourier Transform - STFT) μαζί με την χρήση συναρτήσεων παραθύρου . Η συνάρτηση παράθυρο που χρησιμοποιείται στην δικιά μας περίπτωση είναι η Χάμινγκ (Hamming).Τα χαρακτηριστικά που εξάγουμε είναι τα εξής :

- Τους πρώτους 13 Mel-Frequency Cepstral Coefficients (MFCC) (Davis and Mermelstein, 1990)

- Φασματική επιπεδότητα

  $$Επιπεδότητα = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\frac{\sum_{n=0}^{N-1} x(n)}{N}} \qquad (3)$$

  όπου, x(n) το πλάτος συχνοτικού εύρους n.

- Φασματικό κέντρο (Tjoa, 2017)

  $$Κέντρο = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \qquad (4)$$

  όπου, x(n) το ίδιο με παραπάνω και f(n) η κεντρική συχνότητα του συχνοτικού εύρους n.

- Φασματικό ρόλοφ (Spectral roll of), το οποίο ορίζεται ως η συχνότητα πάνω από την οποία βρίσκεται το 85% της συνολικής φασματικής ενέργειας. (Tjoa, 2017)

- Φασματικό εύρος ζώνης (Tjoa, 2017)

  $$(\sum_k S(k)(f(k) - f_c)^p)^{\frac{1}{p}} \qquad (5)$$

  όπου, S(k) είναι το πλάτος στο συχνοτικό εύρος k, f(k) η κεντρική συχνότητα του εύρους k και f_c το φασματικό κέντρο.

- Φασματικό αντίθεση (Spectral contrast), το οποίο ορίζεται ως η διαφορά της φασματικής κορυφής και της φασματικής κοιλάδας σε 7 φασματικές υπό-λωρίδες (subbands). (Jiang et al., 2002)

Τέλος, εξάγουμε 12 χρωματικά χαρακτηριστικά από το χρωμόγραμμα, το οποίο αναπαριστά την εξάπλωση του τονικού περιεχομένου σε κάθε μια από τις 12 οκτάβες{C, C♯, D, D♯, E , F, F♯, G, G♯, A, A♯, B} κατά τη χρονική διάρκεια του κομματιού. (Ellis, 2007)

2.1.3 Προεπεξεργασία

Στη συνέχεια, ακολουθεί η διαδικασία της προεπεξεργασίας, όπου αναπτύχθηκαν οι παρακάτω τεχνικές:

- Απαλοιφή διπλότυπων χαρακτηριστικών.
- Απαλοιφή χαρακτηριστικών που παρουσιάζουν ελάχιστη διακύμανση.
- Μελέτη συσχέτισης μεταξύ των χαρακτηριστικών, με βάση τον συντελεστή Pearson για τις γραμμικές συσχετίσεις

  $\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$ (6)

  όπου cov η συνδιακύμανση των τυχαίων μεταβλητών X,Y, σ_X η τυπική απόκλιση της X και σ_Y η τυπική απόκλιση της Y.
- Μελέτη συσχέτισης μεταξύ των χαρακτηριστικών, με βάση τον συντελεστή Kendall για τις μη γραμμικές συσχετίσεις.

  $\tau = \frac{2}{n(n-1)} \sum_{i<j} sgn(x_i - x_j) sgn(y_i - y_j)$ (7)

  όπου, $-1 \leq \tau \leq 1$ με την μέγιστη τιμή να λαμβάνεται όταν υπάρχει απόλυτη συσχέτιση μεταξύ 2 χαρακτηριστικών.
- Μελέτη και αξιολόγηση των χαρακτηριστικών με βάση την αμοιβαία πληροφορία αυτών ως προς το χαρακτηριστικό κλάση.

2.1.4 Επιλογή/Μετασχηματισμός χαρακτηριστικών

Συνεχίζοντας, ακολουθεί η διαδικασία επιλογής χαρακτηριστικών, όπου αναπτύχθηκαν οι ακόλουθες τεχνικές:

- Προοδευτική επιλογή χαρακτηριστικών, που αποτελεί μια επαναληπτική διαδικασία κατά την οποία σε κάθε επανάληψη αξιολογούνται τα χαρακτηριστικά, ξεκινώντας από το καθένα ξεχωριστά και προσθέτοντας ένα χαρακτηριστικό ανά επανάληψη, έως ένα σύνολο χαρακτηριστικών του οποίου την πληθικότητα ορίζουμε εμείς, βάσει της μετρικής ακρίβεια, η οποία θα αναλυθεί παρακάτω.

- Οπισθόδρομη επιλογή χαρακτηριστικών, η οποία ακολουθεί την διαδικασία που αναλύθηκε παραπάνω, με την διαφορά ότι το πρώτο σύνολο αποτελούν όλα τα δεδομένα, με αποτέλεσμα να χρησιμοποιείται η οπισθόδρομη διαδικασία, αφαιρώντας ένα χαρακτηριστικό ανά επανάληψη, έως ότου να καταλήξουμε σε ένα σύνολο μικρότερης πληθικότητας.
- Εξαντλητική επιλογή χαρακτηριστικών, όπου αξιολογούνται όλοι οι πιθανοί συνδυασμοί χαρακτηριστικών, σε ένα εύρος πληθικότητας συνόλων που ορίζουμε εμείς, με αποτέλεσμα να επιλέγεται το καλύτερο δυνατό σύνολο.

Για την διαδικασία μετασχηματισμού χαρακτηριστικών και κατ' επέκταση δημιουργία ενός καινούργιου διανύσματος χαρακτηριστικών, με σκοπό την επίτευξη μιας πληρέστερης και συνεκτικής αναπαράστασης των χαρακτηριστικών, υλοποιήθηκαν οι εξής τεχνικές:

- Ανάλυση κύριων συνιστωσών (Principal Component Analysis – PCA), η οποία έχει ως στόχο την δημιουργία ενός μετασχηματισμένου διανύσματος χαρακτηριστικών με βάσει την μεγιστοποίηση της συμμεταβλητότητας μεταξύ αυτών.

- Γραμμική Ανάλυση συνιστωσών (Linear Discriminant Analysis - LDA), η οποία εκτός από την μεγιστοποίηση της συμμεταβλητότητας, μεγιστοποιεί και τις αποστάσεις μεταξύ των κλάσεων.

- Αυτοκωδικοποιητής (Autoencoder), δηλαδή χρήση ενός τεχνητού νευρωνικού δικτύου για την συμπίεση και αποσυμπίεση των χαρακτηριστικών. Έπειτα, χρησιμοποιώντας το κομμάτι του κωδικοποιητή από το δίκτυο μετασχηματίζουμε τα χαρακτηριστικά σε μια νέα συμπαγή μορφή που είναι ικανή να αναπαράξει την αρχική. Επίσης, σε διαφορά με τους παραπάνω γραμμικούς μετασχηματιστές, υπάρχει η δυνατότητα, λόγω της πολυπλοκότητας του νευρωνικού δικτύου, να μοντελοποίησουμε επίσης μη-γραμμικές σχέσεις μεταξύ των χαρακτηριστικών μας.

2.1.5 Ταξινόμηση

Το τελικό στάδιο του μοντέλου μας, το οποίο αποτελεί και αυτό στο οποίο γίνεται η αναγνώριση του είδους είναι η ταξινόμηση. Πιο αναλυτικά, εισάγεται ένα διάνυσμα χαρακτηριστικών στο οποίο καταλήγουμε από τις διάφορες διαδικασίες που αναφέρθηκαν παραπάνω σε διάφορους ταξινομητές της επιλογής μας. Αυτοί οι ταξινομητές είναι η εξής:

- Τυχαίο Δάσος (Random Forest), το οποίο αποτελείται από επιμέρους δέντρα απόφασης που συνδυάζουν τις αποφάσεις τους.
- Διανυσματικές Μηχανές Στήριξης (SVM), που διαχωρίζει με ένα υπερεπίπεδο τα μη-γραμμικά δεδομένα σε έναν άλλο χώρο, στον οποίο τα δεδομένα είναι γραμμικά διαχωρίσιμα. (Cortes and Vapnik, 1995)
- Τεχνητό Νευρωνικό Δίκτυο (ANN), το οποίο αναλόγως τον αριθμό των επιπέδων και τον κόμβων που επιλέγεται, μοντελοποιεί και αντίστοιχη πολυπλοκότητα.

Οι μετρικές αξιολόγησης που χρησιμοποιήθηκαν είναι η ακρίβεια και ο πίνακας σύγχυσης, ο οποίος μας προσφέρει μια πιο λεπτομερή αξιολόγηση της ταξινόμησής μας για κάθε κλάση.

Ένα στοιχείο το οποίο κρίνεται σημαντικό στο πρόβλημα μας είναι ότι υπάρχουν κλάσεις οι οποίες δεν έχουν σαφή όρια με τις υπόλοιπες, όπως για παράδειγμα η ροκ. Αυτό είναι λογικό, μιας και το είδος ροκ από μόνο του αποτελεί ένα ευρές και ασαφές είδος μουσικής.

2.2 Τεχνική χρονικής συνάθροισης χαρακτηριστικών

Η δεύτερη προσέγγιση του προβλήματος,ανεξάρτητη παντελώς από την πρώτη προσέγγιση, αποτελεί την χρονική συνάθροιση των χαρακτηριστικών. Αρχικά χρησιμοποιείται η πρώιμη χρονική συνάθροιση αυτών. Πιο αναλυτικά, οι τιμές των χαρακτηριστικών υπολογίζονται σε μικρά χρονικά πλαίσια και συναθροίζοντας ένα προκαθορισμένο αριθμό τέτοιων πλαισίων δημιουργείται ένα χρονικό παράθυρο υφής.

Έστω Z[k] αποτελεί ένα διάνυσμα U χαρακτηριστικών για τον k-ιοστό χρονικό πλαίσιο, όπου Z[k] = $[z_1[k], z_2[k], \dots, z_u[k]]$. Εφαρμόζοντας Q συναρτήσεις συνάθροισης σε κάθε χαρακτηριστικό z_i στο n-ιοστό χρονικό πλαίσιο, και επεκτείνοντας τον αριθμό των χρονικών πλαισίων από k-L+1 έως k, δημιουργείται ένα διάνυσμα UxQ διαστάσεων, W[n] = $[w_1[n], w_2[n], \dots, w_{UXQ}[n]]$.

Απομονώνοντας ένα συγκεκριμένο χαρακτηριστικό x του αρχικού διανύσματος Z[k] και εφαρμόζοντας μια συνάρτηση f για χρονική συνάθροιση, δημιουργείται ένα X_F διάνυσμα ως εξής:

$$X_F = f(x[k-L+1], \dots, x[k]) \quad (8)$$

Στη θέση της f χρησιμοποιήθηκαν οι παρακάτω στατιστικές μετρικές:

- Μέση τιμή

  $$X_{mean}[n] = \frac{1}{L}\sum_{m=k-L+1}^{k} x[m] \quad (9)$$

- Τυπική απόκλιση

  $$X_{STD}(n) = \frac{1}{L}\sum_{m=k-L+1}^{k} x[m] \quad (10)$$

- Σχετική τυπική απόκλιση

  $$X_{CV} = \frac{X_{MEAN}[n]}{X_{STD}[n]} \quad (11)$$

- Παράγοντας υψηλής κορυφής (High crest factor)

  $$X_{HCF}[n] = \frac{MAX(x[k-L+1],\dots,x[k])}{X_{MEAN}[n]} \quad (12)$$

- Παράγοντας χαμηλής κορυφής (Low crest factor)

  $$X_{LCF}[n] = \frac{MIN(x[k-L+1],\dots,x[k])}{X_{MEAN}[n]} \quad (13)$$

Με τη συγκεκριμένη τεχνική, εισάγονται στους ταξινομητές δεδομένα που εμπεριέχουν πληροφορία για την χρονική εξέλιξη της χρονοσειράς/σήματος και οι διαφορετικές στατιστικές μοντελοποιούν περισσότερες συμπεριφορές αυτής, ενώ παράλληλα μειώνεται και η πολυπλοκότητα των ταξινομητών. (Anders Meng et al., 2007)

Κεφάλαιο 3

3. Βιβλιογραφική επισκόπηση

Η αναγνώριση μουσικού είδους αποτελεί μια περιοχή έρευνας με μεγάλο ενδιαφέρον και επεκτάσεις. Αρχικά, οι Tzanetakis και Cook (2002) μελέτησαν το πρόβλημα με προσεγγίσεις μηχανικής μάθησης, όπως μοντέλα γκαουσιανού μείγματος και κ-κοντινότερων γειτόνων. Επίσης, παρουσίασαν 3 κατηγορίες χαρακτηριστικών για το συγκεκριμένο πρόβλημα: Χαρακτηριστικά χροιάς, ρυθμικού περιεχομένου και τονικού ύψους(pitch content).

Κρυφά Μαρκοβιανά Μοντέλα, που αποτελούν και ταξινομητές όψιμης συνάθροισης, έχουν χρησιμοποιηθεί εκτενώς σε προβλήματα αναγνώρισης ομιλίας και επίσης έχουν διερευνηθεί για προβλήματα αναγνώρισης μουσικού είδους (Scaringella and Zoia, 2005, Soltau et al., 1998).

Μηχανές διανυσματικές στήριξης με διαφορετικές μετρικές απόστασης έχουν χρησιμοποιηθεί για αναγνώριση είδους (Mandel and Ellis, 2005). Οι Lidy and Rauber (2005), μελέτησαν την συμβολή των ψυχό-ακουστικών χαρακτηριστικών για αναγνώριση μουσικού είδους. Χαρακτηριστικά στο πεδίο της συχνότητας, όπως MFCC , φασματική αντίθεση και φασματικό ρόλοφ (spectral rollof) είναι επίσης χαρακτηριστικά που χρησιμοποιούν οι Tzanetakis και Cook, 2002 . Επίσης ένας συνδυασμός ακουστικών και οπτικών χαρακτηριστικών έχουν χρησιμοποιηθεί για εκπαίδευση μηχανών διανυσματικών στήριξης (SVM) και AdaBoost ταξινομητών, Nanni et al. (2016).

Οι Anders Meng et al. (2007) διερεύνησαν την τεχνική της χρονικής συνάθροισης για αναγνώριση μουσικού είδους και υλοποίησαν επίσης μοντέλα αυτο-παλινδρόμησης για πρώιμη συνάθροιση.Οι Lazaros Vrysis et al. (2017) υλοποίησαν επίσης περισσότερες στατιστικές μετρικές για την πρώιμη συνάθροιση με σκοπό την εξαγωγή καλύτερων χαρακτηριστικών.

Κεφάλαιο 4

4. Μεθοδολογία

4.1 Μεθοδολογία πρώτης προσέγγισης

Αρχικά στην πρώτη προσέγγιση χρησιμοποιήθηκαν πλαίσια μήκους 2048 δειγμάτων με ρυθμό δειγματοληψίας 22050 Hz, άρα η χρονική διάρκεια των πλαισιών ανέρχεται στα περίπου 92 ms και επιλέγοντας μήκος άλματος 512 στα μη-επικαλυπτόμενα πλαίσια , στην ουσία υπολογίζονται τα χαρακτηριστικά σε πλαίσια χρονικής διάρκειας 23 ms. Στη συνέχεια, αφού υπολογίζονται οι τιμές των παραπάνω χαρακτηριστικών που αναφέρθηκαν σε κάθε πλαίσιο, συναθροίζονται αυτές οι τιμές κατα μήκος όλων των πλαισίων σε ολόκληρο το κομμάτι των 30 δευτερολέπτων με τη χρήση δύο στατιστικών: Μέση τιμή και τυπική απόκλιση. Κατα αυτόν τον τρόπο σχηματίζεται ένα διάνυσμα 81 χαρακτηριστικών. Παρατηρείται ότι το μήκος του διανύσματος είναι σχετικά μεγάλο , με αποτέλεσμα να προστείθεται πολυπλοκότητα στον ταξινομητή. Για την επιλύση αυτού του προβλήματος , καταφεύγουμε στις μεθοδούς επιλογής και μετασχηματισμού χαρακτηριστικών, που αναφέρθηκαν στο κεφάλαιο δύο, με σκοπό την απομόνωση της χρήσιμης πληροφορίας.

Κατά τη διαδικασία της προεπεξεργασίας, παρατηρήθηκαν χαρακτηριστικά που παρουσίαζαν χαμηλή διακύμανση, αφαιρέθηκαν από το διάνυσμα χαρακτηριστικών και είναι τα παρακάτω:

- Μέση τιμή φασματικής επιπεδότητας
- Τυπική απόκλιση φασματικής επιπεδότητας
- Τυπική απόκλιση του ρυθμού μετάβασης από το σημείο (0,0)

Κατά τη διαδικασία του μετασχηματισμού χαρακτηριστικών, παρατηρώντας από τον πίνακα συσχέτισης μεταξύ αυτών την ύπαρξη μη γραμμικών συσχετίσεων, χρησιμοποιήθηκε αυτοκωδικοποιητής για την καλύτερη μοντελοποίηση αυτών των συσχετίσεων, πέρα από την ανάλυση κύριων συνιστωσών και την γραμμική ανάλυση συνιστωσών που μοντελοποιούν γραμμικές συσχετίσεις.

Από τις διαφορετικές μεθόδους μετασχηματισμού χαρακτηριστικών, το διάνυσμα χαρακτηριστικών μετά την γραμμική ανάλυση συνιστωσών έχει μήκος 9 χαρακτηριστικά, μετά την ανάλυση κύριως συνιστωσών μήκος 32 χαρακτηστικά κρατώντας το 98% της πληροφορίας και μετά τον αυτοκωδικοποιητή μήκος 45 χαρακτηριστικά. Κατά τη διακασία επιλογής χαρακτηριστικών ακολουθώντας τις μεθόδους που αναλύθηκαν στο κεφάλαιο 2 καταλήγουμε επίσης σε διαφορετικό μήκος διανύσματος χαρακτηριστικών.

Το κάθενα διάνυσμα ξεχωριστά παρέχεται ως είσοδος στους ταξινομητές για τη σύγκριση αποτελεσμάτων μεταξύ αυτών.

4.2 Μεθοδολογία δεύτερης προσέγγισης

Στην δεύτερη προσέγγιση, όπου αναφέρθηκε ότι είναι ανεξάρτητη και διαφορετική της πρώτης, διατηρείται χρονική διάρκεια πλαισίου για υπολογισμό χαρακτηριστικών στα 23ms και συναθροίζεται το παράθυρο υφής, όπως αναλύθηκε και στο κεφάλαιο 2 με επιλογή 61 τέτοιων πλαισίων, άρα εν τέλει η χρονική διάρκεια παραθύρου ανέρχεται στα 1,4 δευτερελόπτα.

Οι στατιστικές μέση τιμή, τυπική απόκλιση, σχετική τυπική απόκλιση, παράγοντες υψηλής και χαμηλής κορυφής εφαρμόζονται στα εξής χαρακτηριστικά:

- Ρυθμός μετάβασης από το σημείο(0,0)
- Τετραγωνική ρίζα μέσης ενέργειας του σήματος
- Φασματικό κέντρο
- Φασματικό ρόλοφ (spectral rolloff)
- Φασματική επιπεδότητα
- Φασματικό εύρος

ενώ η μέση τιμή και η τυπική απόκλιση εφαρμόζεται στα MFCCs και στην φασματικη αντίθεση (spectral contrast).

Τα παραπάνω εφαρμόζονται σε καθένα από τα παράθυρα υφής, με αποτέλεσμα να δημιουργείται ένα διάνυσμα χαρακτηριστικών μήκους 69 για κάθε παράθυρο.

Έπειτα, το κάθε διάνυσμα τροφοδοτείται σε μηχανές διανυσματικής στήριξης (SVM), όπου και παίρνεται μια απόφαση για κάθε παράθυρο ξεχωριστά και η τελική απόφαση προκύπτει από ψηφοφορία των επιμέρους αποφάσεων.

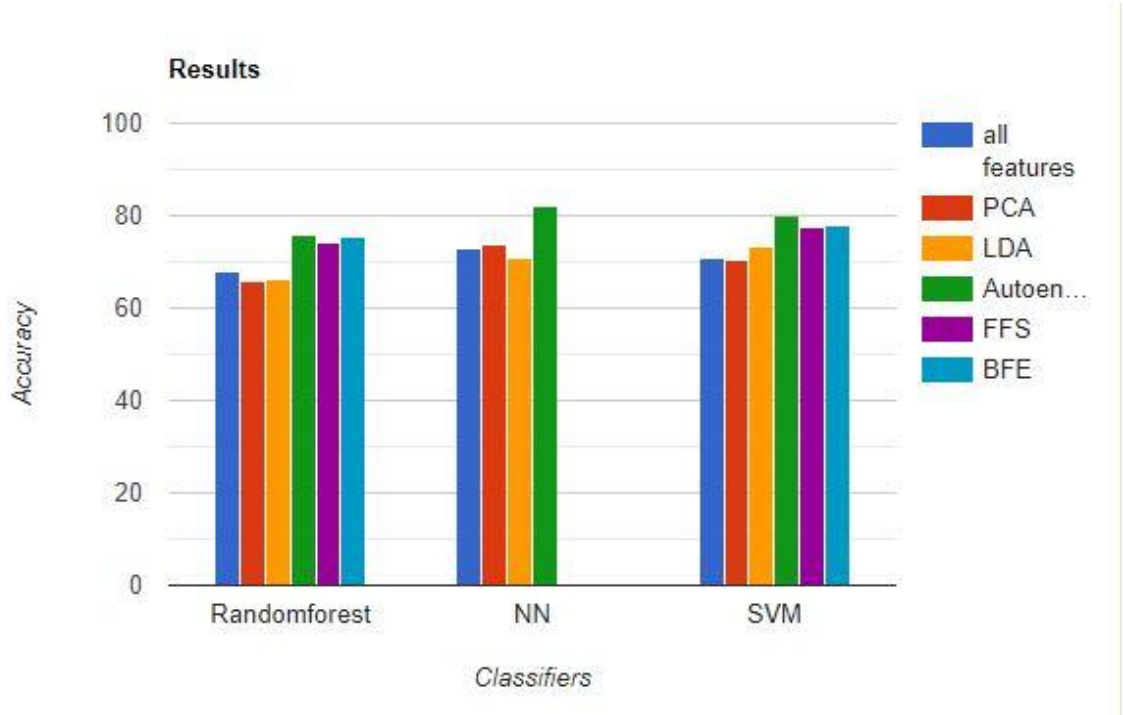
Κεφάλαιο 5

5. Αποτελέσματα

5.1 Αποτελέσματα προσέγγισης με παραδοσιακές τεχνικές μηχανικής μάθησης

Παρακάτω, παρουσιάζονται αναλυτικά τα αποτελέσματα που προέκυψαν ακολουθώντας παραδοσιακές τεχνικές μηχανικής μάθησης.

- Το τυχαίο δάσος παρουσιάζει μεγαλύτερη ακρίβεια (77%), εφόσον έχει προηγηθεί μετασχηματισμός χαρακτηριστικών με τη χρήση αυτοκωδικοποιητή και την ελάχιστη (65%), όταν προηγείται ανάλυση κυρίων συνιστωσών.
- Το τεχνητό νευρωνικό δίκτυο παρουσιάζει την μέγιστη ακρίβεια (82%), εφόσον έχει προηγηθεί μετασχηματισμός χαρακτηριστικών με τη χρήση αυτοκωδικοποιητή και μικρότερη (71%), όταν προηγείται γραμμική ανάλυση συνιστωσών.
- Οι μηχανές διανυσματικής στήριξης παρουσιάζουν μεγαλύτερη ακρίβεια (80%), εφόσον έχει προηγηθεί χρήση αυτοκωδικοποιητή και μικρότερη (70%), με τη χρήση ανάλυσης κυρίων συνιστωσών.
  Στο σχήμα 1, φαίνονται αναλυτικά τα αποτελέσματα, όπου ως FFS ονομάζεται η προοδευτική επιλογή χαρακτηριστικών και BFS η οπισθόδρομη.

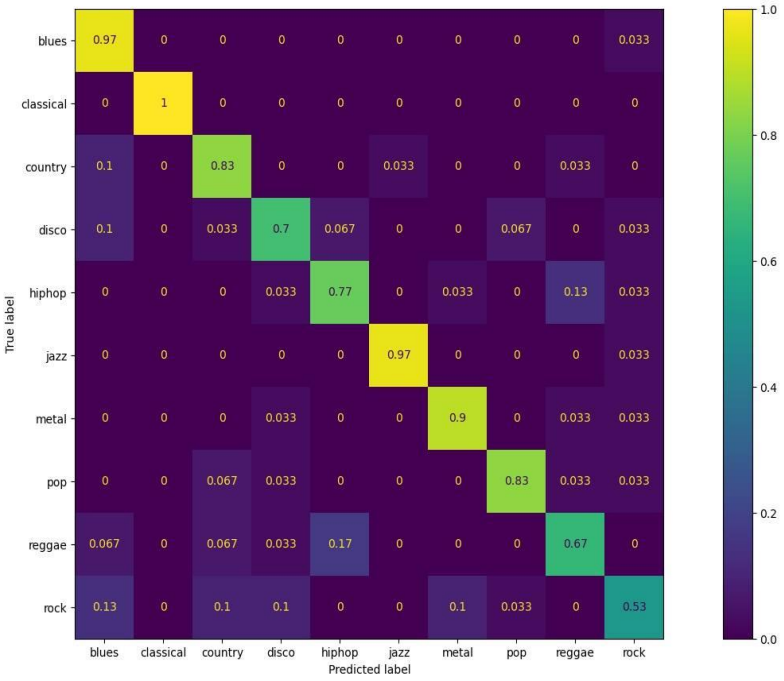


Σχήμα 1: Αποτελέσμα των ταξινομητών με χρήση διάφορων μεθολογιών.

5.2 Αποτελέσματα προσέγγισης με τεχνική χρονικής συνάθροισης χαρακτηριστικών

Παρακάτω, παρουσιάζεται ο πίνακας σύγχυσης που προέκυψε ακολουθώντας την τεχνική χρονικής συνάθροισης χαρακτηριστικών, όπου εξάγονται τα εξής χρήσιμα συμπεράσματα:

- «Τέλεια» ταξινόμηση των μουσικών κομματιών που ανήκουν στο είδος της κλασσικής ,μπλουζ , τζαζ και μεταλ μουσικής.
- Δυσκολία επιτυχούς ταξινόμησης του μουσικού είδους «ροκ», λόγω του γεγονότος ότι αποτελεί ένα ευρές και ασαφές είδος μουσικής.



Σχήμα 2: Πίνακας σύγχυσης των αποτελεσμάτων της δεύτερης προσέγγισης.

Κεφάλαιο 6

6. Συμπεράσματα και μελλοντικές προεκτάσεις

6.1 Συμπεράσματα

Με βάση τα παραπάνω, με τη χρησιμοποίηση παραδοσιακών τεχνικών μας δίνεται η δυνατότητα εξαγωγής,μελέτης και ανάλυσης των χαρακτηριστικών “χειροκίνητα”, ενώ επιτυγχάνονται ικανοποιητικά αποτελέσματα, που καθιστούν τη συγκεκριμένη προσέγγιση πρακτική.

Ακόμη, σε περίπτωση χρησιμοποίησης βαθειάς μάθησης θα υπήρχε μεγαλύτερη υπολογιστική πολυπλοκότητα, ενώ παράλληλα θα αγνοούσαμε την επίδραση των διαφόρων χαρακτηριστικών.

6.2 Μελλοντικές προεκτάσεις

Ένα αρχικό βήμα βελτιστοποιήσης της προτεινούμενης προσέγγισης είναι η χρήση διαφορετικών συνόλων δεδομένων, τα οποία θα περιέχουν μεγαλύτερο αριθμό μουσικών κομματιών και περισσότερα μουσική είδη για την ταξινόμηση αυτών. Πιο συγκεκριμένα, παραδείγματα τέτοιων συνόλων δεδομένων αποτελούν τα παρακάτω:

- https://research.google.com/audioset/ontology/index.html
- http://millionsongdataset.com/

Συνεχίζοντας, είναι δυνατόν να χρησιμοποιηθούν μοντέλα αυτοπαλινδρόμησης για πρώιμη συνάθροιση, με σκοπό την καλύτερη μοντελοποιήση των χρονικών μεταβολών των χαρακτηριστικών.

Χρησιμοποίηση περισσότερων στατιστικών για την ανάλυση της χρονοσειράς/σήματος, με σκόπο την αναπαράσταση χρήσιμης πληροφορίας.

Χρησιμοποίηση ταξινομητών όπως Κρυφών μαρκοβιανών μοντέλων, νευρωνικών δικτύα μακροπρόθεσμης μνήμης (LSTM), ανατροφοδοτούμενων νευρωνικών δικτύων (RNN), οι οποίοι μας παρέχουν καλύτερη μοντελοποίηση της όψιμης χρονικής συνάθροισης.

Αναφορές

- Cortes and Vapnik (1995) Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning* 20(3):273–297.
- Davis and Mermelstein (1990) Steven B Davis and Paul Mermelstein. 1990. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. In *Readings in speech recognition*, Elsevier, pages 65–74.

- Ellis (2007) Dan Ellis. 2007. Chroma feature analysis and synthesis. *Resources of Laboratory for the Recognition and Organization of Speech and Audio-LabROSA.*
- Gouyon et al. (2000) Fabien Gouyon, François Pachet, Olivier Delerue, et al. 2000. On the use of zero-crossing rate for an application of classification of percussive sounds. In *Proceedings of the COST G-6 conference on Digital Audio Effects (DAFX-00), Verona, Italy.*
- Grosche et al. (2010) Peter Grosche, Meinard Müller, and Frank Kurth. 2010. Cyclic tempogramâ mid-level tempo representation for musicsignals. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, pages 5522 – 5525.
- Jiang et al. (2002) Dan-Ning Jiang, Lie Lu, Hong-Jiang Zhang, Jian-Hua Tao, and Lian-Hong Cai. 2002. Music type classification by spectral contrast feature. In *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*. IEEE, volume 1, pages 113-116.
- Lidy and Rauber (2005) Thomas Lidy and Andreas Rauber. 2005. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In ISMIR. pages 34-41.
- Mandel and Ellis (2005) Michael I Mandel and Dan Ellis. 2005. Song-level features and support vector machines for music classification. In *ISMIR*. volume 2005, pages 594-599.
- Nanni et al. (2016) Loris Nanni, Yandre MG Costa, Alessandra Lumini, Moo Young Kim, and Seung Ryul Baek. 2016. Combining visual and acoustic features for music genre classification. Expert Systems with Applications 45:108-117.
- Scaringella and Zoia (2005) Nicolas Scaringella and Giorgio Zoia. 2005. On the modeling of time information for automatic genre recognition systems in audio signlas. In ISMIR. pages 666-671.
- Tjoa (2017) Steve Tjoa. 2017. Music information retrieval. https://musicinformationretrieval.com/spectral_features.html. Accessed: 2018-02-20.
- Tzanetakis and Cook (2002) ,George Tzanetakis and Perry Cook. 2002. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing 10(5):293-302.*
- Soltau et al. (1998) Hagen Soltau, Tanja Schultz, Martin Westphal, and Alex Waibel. 1998. Recognition of music types. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*. IEEE, volume 2, pages 1137–1140.
- Anders Mengs et al. (2007) Peter Ahrendt, Jan Larsen and Lars Kai Hansen. *Temporal Feature Integration for Music Genre Classification. IEEE Transactions on audio, speech and language processing.*
- Lazaros Vrysis et al. (2017) Nikolaos Tsipas, Charalampos Dimoulas and George Papanikolaou. *Extending Temporal Feature Integration for Semantic Audio Analysis.* In the Journal of the Audio Engineering Society.