

Ατομική Διπλωματική Εργασία

**ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ
ΣΤΗ ΜΟΥΣΙΚΗ**

Αυξέντης Αυξέντη

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μάιος 2014

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Υπολογιστική Νοημοσύνη στη Μουσική

Αυξέντης Αυξέντη

Επιβλέπων Καθηγητής

Χρίστος Ν. Σχίζας

Η Ατομική Διπλωματική Εργασία υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων απόκτησης του πτυχίου Πληροφορικής του Τμήματος Πληροφορικής του Πανεπιστημίου Κύπρου

Μάιος 2014

Ευχαριστίες

Θα ήθελα να ευχαριστήσω όλους εκείνους που με τον δικό τους τρόπο συντέλεσαν στην ολοκλήρωση αυτής της Ατομικής Διπλωματικής Εργασίας. Ιδιαίτερες ευχαριστίες οφείλω στον επιβλέποντα Καθηγητή Χρίστο Ν. Σχίζα που από την πρώτη στιγμή με εμπιστεύθηκε για την κατανόηση, τη σωστή καθοδήγηση και για την διεκπεραίωση της παρούσας εργασίας.

Ένα μεγάλο ευχαριστώ στον Ανδρέα Νεοκλέους για της συμβουλές, την καθοδήγηση την υπομονή και τις ιδέες που μου έχει προσφέρει για την ολοκλήρωση της Ατομικής Διπλωματικής Εργασίας.

Τέλος των ευχαριστιών μου δεν θα μπορούσα να παραλείψω την οικογένεια μου, τους φίλους μου, τον Μάριο Ανθίμου και όσους ήταν αυτό το χρονικό διάστημα στο πλάι μου καθώς επίσης για την υποστήριξη και την βοήθεια που μου πρόσφεραν.

Περίληψη

Στόχος της παρούσας Ατομικής Διπλωματικής Εργασίας είναι η προσπάθεια δημιουργίας ενός αλγόριθμου που να αναγνωρίζει επαναλήψεις στη μελωδία ενός τραγουδιού.

Καταρχάς, έγινε μια εισαγωγή στην εθνομουσικολογία και παρουσιάστηκαν οι στόχοι της Ατομικής Διπλωματικής Εργασίας. Έπειτα, περιγράψαμε το πρόβλημα για το παρόν σύστημα που αναλύσαμε και αναφέραμε άλλα παρόμοια συστήματα. Ακολούθως, επισημάναμε την απαιτούμενη γνώση και τεχνολογία που χρειάστηκε για να φέρουμε εις πέρας αυτή την εργασία, όπως MATLAB, Ψηφιακή Επεξεργασία Σήματος, Weka κτλ. Για την ανάλυση των απαιτήσεων του συστήματος, έγινε βιβλιογραφική κάλυψη διαφόρων συναρτήσεων της MATLAB καθώς επίσης και έτοιμων συναρτήσεων από προηγούμενη δουλειά, με σκοπό των καθορισμό των προδιαγραφών του συστήματος.

Για την υλοποίηση αυτής της Ατομικής Διπλωματικής Εργασίας, χρησιμοποιήθηκε μια βάση δεδομένων με 9 μονοφωνικά λαϊκά τραγούδια της Κύπρου για την επικύρωση των μεθόδων COSFIRE και DTW. Σε κάθε τραγούδι εφαρμόστηκε μια συγκεκριμένη διαδικασία για την εξαγωγή κάποιων φράσεων με σκοπό να ελεγχθούν στη συνέχεια αν είναι όμοιες.

Το σύστημα μας παρουσίασε μια σημαντική διαφορά μεταξύ των δύο μεθόδων COSFIRE και DTW, που είναι ο χρόνος που χρειάζεται για να εκτελέσει ένα συγκεκριμένο τραγούδι ο κάθε αλγόριθμος. Έχει αποδειχθεί ότι ο χρόνος της επεξεργασίας του DTW είναι σημαντικά υψηλότερος από τη μέθοδο COSFIRE. Εν κατακλείδι, καταλήγουμε στο συμπέρασμα ότι η μέθοδος COSFIRE αποδίδει καλύτερα από τη μέθοδο DTW.

Υπεύθυνα δηλώνω ότι το περιεχόμενο αυτής της εργασίας δεν περιέχει καμία αντιγραφή από άλλη πηγή χωρίς να γίνεται η νενομισμένη αναφορά στην πηγή. Έχω ελέγξει την εργασία μου με πρόγραμμα λογοκλοπής.

Περιεχόμενα

Κεφάλαιο 1	Εισαγωγή.....	1
	1.1 Εισαγωγή στην Εθνομουσικολογία	1
	1.2 Στόχοι της Ατομικής Διπλωματικής Εργασίας	2
Κεφάλαιο 2	Περιγραφή προβλήματος.....	3
	2.1 Περιγραφή βασικών εννοιών του υπομελέτη προβλήματος	3
	2.2 Άλλα παρόμοια Συστήματα	4
Κεφάλαιο 3	Απαιτούμενη Γνώση και Τεχνολογίες.....	7
	3.1 Ανάπτυξη απαιτούμενων γνώσεων και τεχνολογιών	7
	3.1.1 NeuroShell	7
	3.1.2 Matlab	8
	3.1.3 Weka	9
	3.1.4 Ψηφιακή Επεξεργασία Σήματος	10
	3.1.5 Τεχνικές Μηχανικής Μάθησης	10
	3.1.5.1 Μάθηση με Επίβλεψη	11
	3.1.5.2 Μάθηση χωρίς Επίβλεψη	13
	3.1.5.3 Άλλα Είδη Μάθησης	14
Κεφάλαιο 4	Ανάλυση Απαιτήσεων, Προδιαγραφές.....	15
	4.1 Συναρτήσεις Matlab	15
Κεφάλαιο 5	Σχεδιασμός Συστήματος, Υλοποίηση	21
	5.1 Μεθοδολογία	21
	5.2 Ο αλγόριθμος YIN	23
	5.3 Προ-Επεξεργασία	24
	5.4 Σχεδιασμός φίλτρων COSFIRE	27
	5.5 Δυναμική Στρέβλωση Χρόνου (Dynamic Time Warping - DTW)	30

5.6 Επίτευξη ανοχής συχνότητας	30
Κεφάλαιο 6 Αξιολόγηση Συστήματος – Αποτελέσματα – Συζήτηση.....	31
6.1 Αξιολόγηση Συστήματος – Αποτελέσματα	31
Κεφάλαιο 7 Αποτελέσματα και Μελλοντική Εργασία.....	34
7.1 Συμπεράσματα	34
7.2 Μελλοντική Εργασία	35
Βιβλιογραφία	37

Κεφάλαιο 1

ΕΙΣΑΓΩΓΗ

1.1	Εισαγωγή στην Εθνομουσικολογία	1
1.2	Στόχοι της Ατομικής Διπλωματικής Εργασίας	2

1.1 Εισαγωγή στην Εθνομουσικολογία

Εθνομουσικολογία [1] είναι η επιστήμη που μελετά τη μουσική στα πλαίσια του πολιτισμού και ως πολιτισμό. Οι απαρχές της εθνομουσικολογίας βρίσκονται στα κείμενα των περιηγητών του δεκάτου ογδόου και δεκάτου ενάτου αιώνα, στα οποία συναντάμε περιγραφές των μουσικών, που συναντούσαν οι περιηγητές αυτοί στα ταξίδια τους ανά τον κόσμο.

Η εθνομουσικολογία αναπτύσσεται ως επιστήμη με τη μορφή της συγκριτικής μουσικολογίας στον γερμανικό χώρο στις αρχές του εικοστού αιώνα και την έμφασή της στη μουσική διάσταση παράλληλα με την ανάπτυξη διαφόρων εθνικών μουσικολογικών σχολών ανά την Ευρώπη.

Το «κύριο» καθήκον της Εθνομουσικολογίας είναι να εξηγήσει τι είναι μουσική σε σχέση με το κοινωνικό αλλά από την άποψη των μουσικών παραγόντων. Η έρευνα της Εθνομουσικολογία δηλαδή η ανάλυση του ήχου γίνεται «με το αυτί» από εκπαιδευμένους επιστήμονες.

Κατά την εποχή του Chales Seeger έχουν γίνει πάρα πολλές προσπάθειες για διευκόλυνση της ανάλυσης χρησιμοποιώντας διάφορα τεχνολογικά εργαλεία. Μετά από πάρα πολλές έρευνες έχουν ανοίξει αρκετές γραμμές, αρκετοί δρόμοι για τους ανθρώπους που πραγματικά ενδιαφέρονται σε αυτόν τον τομέα.

Στην έρευνα της Εθνομουσικολογίας μετά από ερευνητικές προσπάθειες θα χρησιμοποιήσουμε τον «Όρο Υπολογιστική Εθνομουσικολογία» όπου αυτός ο όρος

αναφέρετε στον σχεδιασμό, την ανάπτυξη και την χρήση εργαλείων του Ηλεκτρονικού Υπολογιστή τα οποία έχουν το δυναμικό να βοηθήσουν στην έρευνα αυτή.

Μια νέα περιοχή που μπορεί να μας βοηθήσει στην έρευνα αυτή είναι το MIR[2](Music Information Retrieval Μουσική Ανάκτηση Πληροφοριών). Αυτή η περιοχή έχει να κάνει με τον σχεδιασμό, την κατανόηση και την κατασκευή εργαλείων τα οποία θα βοηθήσουν στην οργάνωση, κατανόηση και στην αναζήτηση μεγάλων συλλογών απ μουσική καθώς τα τελευταία χρόνια και αυτό το πεδίο, περιοχή έχει ραγδαία εξέλιξη. Κυρίως το MIR έχει επικεντρωθεί είτε στην Λαϊκή μουσική είτε στην Δυτική «κλασικής» μουσικής.

1.2 Στόχοι της Ατομικής Διπλωματικής Εργασίας

- Ανασκόπηση της βιβλιογραφίας (ανάλυση της μουσικής με υπολογιστικές μεθόδους όπως η τεχνικές μηχανικής μάθησης, εξαγωγή χαρακτηριστικών κ.τ.λ.
- Υλοποίηση διάφορων τεχνικών με την πλατφόρμα matlab.
- Προεπεξεργασία των δεδομένων με σκοπό να κατανοηθεί καλύτερα η βάση δεδομένων και να γίνει πιο αποδοτική για το σύστημα που θα υλοποιήσουμε.
- Να δημιουργηθεί πλατφόρμα στο διαδίκτυο όπου συμπεριλαμβάνονται όλα τα σχετικά της Ατομικής Διπλωματικής Εργασίας.
- Συναντήσεις με τον καθηγητή σε τακτικά χρονικά διαστήματα με σκοπό για την ενημέρωση της δουλείας και να δέχομαι τις παρατηρήσεις του.
- Να τελειώσω στην ώρα μου σύμφωνα με το χρονοδιάγραμμα.

Κεφάλαιο 2

Περιγραφή προβλήματος

2.1 Περιγραφή βασικών εννοιών του υπομελέτη προβλήματος	3
2.2 Άλλα παρόμοια Συστήματα	4

2.1 Περιγραφή βασικών εννοιών του υπό μελέτη προβλήματος

Τα δημοτικά τραγούδια συνήθως αποτελούνται από επαναλήψεις παρόμοιων μελωδικών στροφών. Ένα από τα μεγαλύτερα ενδιαφέροντα στην επιστήμη της μουσικολογίας είναι ο εντοπισμός της δομής ενός τραγουδιού. Από την γέννηση της μουσικολογίας οι επιστήμονες ανέλυναν τα τραγούδια ένα προς ένα με αποτέλεσμα να χρειάζεται πολύς χρόνος για να μελετηθεί η μουσική κουλτούρα ή διάφορα είδη μουσικής. Η επιστήμη της πληροφορικής έδωσε μια άλλη προσέγγιση στην ανάλυση της μουσικής χρησιμοποιώντας μαθηματικά μοντέλα και αλγόριθμους. Αυτό, δημιούργησε μια ώθηση της μουσικολογικής κοινότητας προς σε αυτή την κατεύθυνση.

Το θέμα που θα με απασχολήσει κατά την διάρκεια της έρευνας μου θα είναι η προσπάθεια δημιουργίας ενός αλγόριθμου που να αναγνωρίζει επαναλήψεις στη μελωδία ενός τραγουδιού. Ο αλγόριθμος είναι επηρεασμένος από τον αλγόριθμο COSFIRE [3] όπου με την δημιουργία κάποιων φίλτρων, το σύστημα εκπαιδεύεται για ένα μικρό κομμάτι της μελωδίας και επαληθεύεται στο υπόλοιπο κομμάτι. Το σύστημα δίνει απόκριση μόνο στα σημεία που η μελωδία μοιάζει με το κομμάτι που εκπαιδεύτηκε. Με αυτό τον τρόπο εντοπίζονται οι παρόμοιες μελωδίες στο τραγούδι. Μέσα από μια λογική σειρά, ο αλγόριθμος εντοπίζει τα χρονικά όρια στα οποία εμφανίζονται αυτές οι επαναλήψεις και αυτό ονομάζεται κατάτμηση.

Η κατάτμηση έχει σκοπό να βρει τα όρια μεταξύ των επαναλαμβανόμενων στροφών. Οι περισσότεροι μέθοδοι κατάτμησης ήχου αναπτύχθηκαν για τη pop και

κλασική μουσική, ωστόσο αυτοί δεν αποδίδουν καλά σε ηχογραφήσεις δημοτικών τραγουδιών.

2.2 Άλλα παρόμοια Συστήματα

Η δημοτική μουσική λαμβάνει αυξημένη προσοχή από την κοινότητα MIR (Music Information Retrieval), καθώς η ευαισθητοποίηση μας για τη διατήρηση της πολιτιστικής μας κληρονομιάς και την ανάδειξη της στο ευρύ κοινό μεγαλώνει. Τα δημοτικά τραγούδια συνήθως αποτελούνται από επαναλαμβανόμενα κομμάτια και στροφές. Αυτό είναι τελείως διαφορετικό από την κατάτμηση της pop μουσικής, όπου τα τραγούδια συνήθως αποτελούνται από διάφορα μέρη, όπως η εισαγωγή, η στροφή, η γέφυρα και το ρεφρέν.

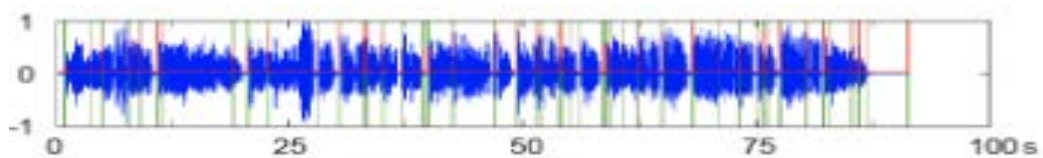
Για να βρούμε τα κομμάτια αυτά σε ηχογραφήσεις δημοτικών τραγουδιών, συνήθως χρησιμοποιούνται μέθοδοι κατάτμησης που χωρίζουν μια εγγραφή σε ξεχωριστά τμήματα σύμφωνα με διαφορετικά κριτήρια. Μια μέθοδος κατάτμησης παίρνει ως είσοδο μια εγγραφή τραγουδιού και για έξοδο μας δίνει ένα σύνολο από όρια, που αντιπροσωπεύουν τα όρια των επιμέρους στροφών στην ηχογράφιση. Στην περίπτωση της ανάλυσης δημοτικών τραγουδιών η μέθοδος θα πρέπει να λαμβάνει υπόψη ότι οι καλλιτέχνες δεν είναι επαγγελματίες τραγουδιστές, η οποία μπορεί να οδηγήσει σε σημαντικές διαφορές στο ρυθμό και την μελωδία.

Οι μέθοδοι υψηλού επιπέδου κατάτμησης μπορούν να γίνουν στο χέρι ή με τη χρήση αυτοματοποιημένων μεθόδων. Για την ακριβή ανάλυση των επιμέρους τραγουδιών, περαιτέρω κατάτμηση σε μικρότερα τμήματα του τραγουδιού είναι επιθυμητή. Αυτό συμβαίνει κυρίως επειδή οι ηχογραφήσεις δημοτικών τραγουδιών περιέχουν μια σειρά από συγκεκριμένα θέματα που αγνοούν αυτές οι μέθοδοι, όπως είναι το ανακριβές τραγούδι των καλλιτεχνών, ο μεταβαλλόμενος ρυθμός σε όλο το τραγούδι και η παρουσία του θορύβου (ομιλίες στο τραγούδι, ο άνεμος και άλλοι περιβαλλοντικοί θόρυβοι, παλαμάκια κτλ.) και οι τραγουδιστές είναι ως επί το πλείστον ανεκπαιδευτοι και συνήθως ηλικιωμένοι που μπορεί να τραγουδούν εκτός ρυθμού, να ξεχνούν τους στίχους ή τη μελωδία, να διακόπτουν τις παραστάσεις τους, να αλλάζουν από τραγούδι σε ομιλία κλπ. Έτσι το πρόβλημα είναι ότι είναι δύσκολο να βρεθεί ένας αλγόριθμος που να εντοπίζει παρόμοιες λειτουργίες δημοτικών τραγουδιών σε ένα τραγούδι, αν και κατά τα

τελευταία έτη αναπτύχθηκαν διάφορες μέθοδοι για την κατάτμηση των δημοτικών τραγουδιών.

Στην μελέτη των Bohak και Marolt [4] παρουσιάζεται μια συγκεκριμένη μεθοδολογία για τον εντοπισμό τέτοιων επαναλήψεων σε παραδοσιακές μουσικές της Σλοβενίας. Η μέθοδος αποτελείται από διάφορα βήματα: προεπεξεργασία, αναζήτηση για φωνητικές παύσεις, αναζήτηση για πιθανές αρχές των στροφών και επιλογή της πραγματικής αρχής των στροφών. Στην προεπεξεργασία, το σήμα ήχου που παίρνουμε ως είσοδο μετατρέπεται από στερεοφωνικό σε ένα μονό κανάλι, η συχνότητα δειγματοληψίας (sample rate) μειώνεται στα 11025 Hz και το πλάτος κανονικοποιείται. Στη συνέχεια, γίνεται αναζήτηση για φωνητικές παύσεις. Οι ερμηνευτές των τραγουδιών είναι συνήθως ερασιτέχνες τραγουδιστές που κάνει χαρακτηριστικές παύσεις για αναπνοή, που αντικατοπτρίζεται σε ηχογραφήσεις σαν τη σιωπή. Αυτές οι παύσεις γίνονται μεταξύ των στροφών, έτσι ώστε να μπορούν να χρησιμοποιηθούν για ανίχνευση των ορίων μεταξύ των στροφών[5]. Χρησιμοποιούνται 3 χαρακτηριστικά για την ανίχνευση των φωνητικών παύσεων:

- 1) Βραχυπρόθεσμη ενέργεια σήματος (short-term signal energy)
- 2) Πλάτος φακέλου του σήματος (amplitude envelope of the signal)
- 3) Ανίχνευση τόνου (detected pitch)



Σχήμα 2.2.1. Υποψήφια όρια στοφών

Μετά γίνεται αναζήτηση για τα όρια των στροφών. Στην αναζήτηση υποψηφίων για τα όρια της στροφής, συγχωνεύουμε όλα τα σύνολα των φωνητικών παύσεων που λαμβάνονται προηγουμένως με τις μεθόδους που περιγράψαμε πιο πάνω. Ένα παράδειγμα ενός τέτοιου συγχωνευμένου κομματιού φαίνεται στο σχήμα 2.2.1, όπου οι αρχές των φωνητικών παύσεων παραλείπονται και μόνο το τέλος τους, το οποίο θεωρούμε ως υποψήφιο για όριο της στροφής, εμφανίζεται με πράσινο χρώμα. Αν οι υποψήφιοι περιλαμβάνονται σε διάφορα σύνολα πριν από τη συγχώνευση, τότε συγχωνεύονται σε ένα μόνο όριο υποψηφίων.

Τέλος, έχουμε την επιλογή της πραγματικής αρχής της στροφής. Η επιλογή των πραγματικών αρχών των στροφών γίνεται με ένα απλό αλγόριθμο που επιλέγει κορυφές σε συνδυασμό με ένα γενικό κατώφλι. Στην συνάρτηση καταλληλότητας, οι κορυφές αντιπροσωπεύουν το πιο πιθανό ξεκίνημα μιας στροφής, έτσι ώστε όλες οι κορυφές πάνω από το γενικό κατώφλι, που αντιστοιχούν στη μέση τιμή της συνάρτησης καταλληλότητας, να επιλέγονται ως τα πραγματικά όρια μεταξύ των στροφών.

Όταν εκτελεστεί η μέθοδος, εξάγονται αποτελέσματα και φθάνουμε σε συμπεράσματα. Η μέθοδος βασίζεται στην ανίχνευση των φωνητικών παύσεων, οι οποίες αντιπροσωπεύουν υποψήφια όρια των στροφών και στη συνέχεια αξιολογούνται σύμφωνα με την ομοιότητα των μελωδιών τους με την πρώτη στροφή. Οι αλγόριθμοι ανίχνευσης φωνητικής παύσης και η μέθοδος στο σύνολό της αξιολογείται ξεχωριστά σε ένα σύνολο δεδομένων από ηχογραφήσεις δημοτικών τραγουδιών. Η μέθοδος λειτουργεί καλά, αφού λειτουργώντας αυτόματα στην ανάλυση μουσικής, διευκολύνει τη δουλειά του μουσικολόγου, έχουν όμως προγραμματιστεί αρκετές επεκτάσεις.

Κεφάλαιο 3

Απαιτούμενη Γνώση και Τεχνολογίες

3.1 Ανάπτυξη απαιτούμενων γνώσεων και τεχνολογιών	7
3.1.1 NeuroShell	7
3.1.2 Matlab	8
3.1.3 Weka	9
3.1.4 Ψηφιακή Επεξεργασία Σήματος	10
3.1.5 Τεχνικές Μηχανικής Μάθησης	10
3.1.5.1 Μάθηση με Επίβλεψη	11
3.1.5.2 Μάθηση χωρίς Επίβλεψη	13
3.1.5.3 Άλλα Είδη Μάθησης	14

3.1 Ανάπτυξη απαιτούμενων γνώσεων και τεχνολογιών

3.1.1 NeuroShell

Το NeuroShell περιλαμβάνει υπερσύγχρονους αλγόριθμους οι οποίοι εκπαιδεύονται γρήγορα, έχοντας την δυνατότητα να επιλύουν αποτελεσματικά τα προβλήματα πρόβλεψης, πρόγνωσης και εκτίμησης σε ένα ελάχιστο ποσοστό του χρόνου, χωρίς να περάσει από την διαδικασία των παραμέτρων του δικτύου. Είναι σχεδιασμένο για να είναι εξαιρετικά εύκολο στη χρήση του. Επίσης διαβάζει και γράφει αρχεία κειμένου για συμβατότητα με πολλά άλλα προγράμματα.

Ο αλγόριθμος πρόβλεψης για να φτάσει στο στάδιο που βρίσκεται σήμερα, πέρασαν αρκετά χρόνια ερευνών. Ο αλγόριθμος πρόβλεψης ενός νευρωνικού δικτύου βοηθά τον εμπειρογνώμονα ή των στατιστικολόγο να κάνει ένα μοντέλο προσαρμόζοντας με ευκολία τις παραμέτρους του καθώς επίσης και στην κατασκευή μοντέλου ανάλογα με τις προβλέψεις μας.

Το NeuroShell είναι τόσο εύκολο στη χρήση του που δεν χρειάζεται εγχειρίδιο. Υπάρχει ένας «Εκπαιδευτής» που μας καθοδηγεί μέσα από την κατασκευή των μοντέλων πρόβλεψης. Σε κάθε στάδιο του εκπαιδευτή, ένα αρχείο βοήθειας μας δίνει όλες τις πληροφορίες που χρειαζόμαστε. Όταν μάθουμε από τον εκπαιδευτή, μπορούμε να το απενεργοποιήσουμε και να εργαστούμε από τη γραμμή εργαλείων ή μενού.

3.1.2 Matlab

Το MATLAB [6] είναι ένα σύγχρονο ολοκληρωμένο μαθηματικό λογισμικό πακέτο που χρησιμοποιείται σε πανεπιστημιακά μαθήματα αλλά και ερευνητικές και άλλες εφαρμογές που ασχολούνται με επιστημονικούς υπολογισμούς (scientific computing). Το MATLAB είναι ένα διαδραστικό (interactive) πρόγραμμα για αριθμητικούς υπολογισμούς και οπτικοποίηση δεδομένων (data visualization) με δυνατότητες προγραμματισμού που το καθιστούν ένα ισχυρό και χρήσιμο εργαλείο στις μαθηματικές και φυσικές επιστήμες. Στις αρχικές του εκδόσεις δεν έκανε συμβολικούς υπολογισμούς.

Στις νεότερες εκδόσεις του, το πακέτο περιλαμβάνει εργαλείοι που επιτρέπουν συμβολικούς υπολογισμούς. Όπως υποδηλώνεται και από το όνομά του, το MATLAB είναι ειδικά σχεδιασμένο για υπολογισμούς με πίνακες, όπως η επίλυση γραμμικών συστημάτων, η εύρεση ιδίων τιμών και ιδίων διανυσμάτων, ή αντιστροφή τετραγωνικών πινάκων κλπ. Επιπλέον το πακέτο αυτό είναι εφοδιασμένο με πολλές επιλογές για γραφικά (δηλ. την κατασκευή γραφικών παραστάσεων) και προγράμματα γραμμένα στη δική του γλώσσα προγραμματισμού για την επίλυση άλλων προβλημάτων όπως η εύρεση των ριζών μη γραμμικής εξίσωσης, η επίλυση μη γραμμικών συστημάτων, η επίλυση προβλημάτων αρχικών τιμών με διαφορικές εξισώσεις κ.α.

Η γλώσσα προγραμματισμού του MATLAB δίνει την ευχέρεια στον χρήστη να το επεκτείνει με δικά του προγράμματα. Συχνά θα γράφουμε η MATLAB (εννοώντας τη γλώσσα προγραμματισμού) και όχι το (πακέτο) MATLAB. Το MATLAB είναι σχεδιασμένο για την αριθμητική επίλυση προβλημάτων σε αριθμητική πεπερασμένη ακρίβειας (finite - precision arithmetic), δηλαδή δεν βρίσκει την ακριβή αλλά μια προσεγγιστική λύση ενός προβλήματος.

3.1.3 Weka

Το Weka [7] περιέχει μια συλλογή από εργαλεία απεικόνισης και αλγορίθμων για την ανάλυση των δεδομένων και την προγνωστική μοντελοποίηση, μαζί με γραφικά περιβάλλοντα χρήστη (GUI) για εύκολη πρόσβαση σε αυτήν τη λειτουργία. Το Weka υποστηρίζει διάφορες λειτουργίες εξόρυξης δεδομένων, πιο συγκεκριμένα, την προεπεξεργασία δεδομένων, ομαδοποίηση, ταξινόμηση, παλινδρόμηση, οπτικοποίηση και επιλογή χαρακτηριστικών.

Όλες οι τεχνικές Weka βασίζονται στην υπόθεση ότι τα δεδομένα είναι διαθέσιμα ως ένα απλό αρχείο ή σχέση, όπου κάθε σημείο δεδομένων περιγράφεται από ένα σταθερό αριθμό από χαρακτηριστικά. Το Weka παρέχει πρόσβαση σε SQL βάσεις δεδομένων που χρησιμοποιούν Java Database Connectivity και μπορεί να επεξεργαστεί το αποτέλεσμα που επιστρέφεται από ένα query βάσης δεδομένων. Δεν είναι σε θέση να πραγματοποιεί εξόρυξη δεδομένων με πολλαπλές σχέσεις μεταξύ τους, αλλά υπάρχει ξεχωριστό λογισμικό για τη μετατροπή μια συλλογής των συνδεδεμένων πινάκων της βάσης δεδομένων σε ένα απλό πίνακα που είναι κατάλληλος για επεξεργασία χρησιμοποιώντας το Weka.

Ένας άλλος σημαντικός τομέας που δεν καλύπτεται επί του παρόντος από τους αλγορίθμους που περιλαμβάνονται στο Weka είναι η μοντελοποίηση ακολουθιών. Η κύρια διεπαφή χρήστη του Weka είναι ο Explorer, αλλά ουσιαστικά η ίδια λειτουργικότητα μπορεί να προσεγγιστεί μέσω της διεπαφής «Knowledge Flow» που είναι βασισμένη σε συστατικά και από τη γραμμή εντολών.

Υπάρχει επίσης, ο πειραματιστής, ο οποίος επιτρέπει τη συστηματική σύγκριση των αλγορίθμων μηχανικής μάθησης του Weka όσον αφορά μια συλλογή συνόλων των δεδομένων.

Πλεονεκτήματα του Weka

- Ελεύθερη διάθεση υπό την άδεια GNU General Public.
- Φορητότητα, δεδομένου ότι εφαρμόζεται πλήρως στη γλώσσα προγραμματισμού Java και ως εκ τούτου λειτουργεί σχεδόν σε κάθε σύγχρονη υπολογιστική πλατφόρμα.
- Μια ολοκληρωμένη συλλογή των δεδομένων προεπεξεργασίας και τεχνικές μοντελοποίησης.
- Ευκολία χρήσης, λόγω γραφικών διεπαφών χρήστη.

3.1.4 Ψηφιακή Επεξεργασία Σήματος

Η ψηφιακή επεξεργασία σήματος [8] ασχολείται με την ψηφιακή αναπαράσταση των σημάτων με την βοήθεια ψηφιακών επεξεργαστών.

Ορισμένα παραδείγματα της ψηφιακής επεξεργασίας σήματος είναι: (α) περιπτώσεις κατά τις οποίες θέλουμε να αφαιρέσουμε τον θόρυβο από ένα σήμα, (β) να βρούμε τον μετασχηματισμό Fourier[9] κάποιων δεδομένων, (γ) να μετατρέψουμε ένα σήμα σε μια μορφή πιο κατάλληλη για επεξεργασία και ανάλυση της πληροφορίας που εμπεριέχει κτλ. Ασχολείται ακόμη με την γραμμικότητα ενός συστήματος στην οποία συνεπάγεται ότι δύο διαφορετικά σήματα μπορούν να διέλθουν μέσα από το σύστημα ταυτοχρόνως χωρίς να επηρεάζουν το ένα το άλλο. Ένα σημαντικό χαρακτηριστικό των γραμμικών συστημάτων είναι ότι αν η είσοδος είναι ένα απλό ημιτονοειδές σήμα, τότε η έξοδος είναι ένα ημίτονο ίδιας συχνότητας αλλά με τροποποιημένο πλάτος και φάση. Η χαμηλότερη αυτή ημιτονοειδής συχνότητα ονομάζεται θεμελιώδης συχνότητα (fundamental frequency).

Ένα ακόμη πλεονέκτημα της ψηφιακής επεξεργασίας ενός σήματος έναντι της αντίστοιχης αναλογικής είναι ότι είναι χαμηλότερου κόστους. Αυτό μπορεί να οφείλεται είτε στο ότι το υλικό (hardware) σήμερα είναι φθηνότερο, είτε στην απλοποίηση που παρέχεται λόγω της ψηφιακής υλοποίησης.

Κάποιες εφαρμογές της ψηφιακής επεξεργασίας σήματος είναι η επεξεργασία ήχου, η αναγνώριση φωνής, η επεξεργασία σημάτων από ραντάρ, η στατιστική επεξεργασία σήματος, η ψηφιακή επεξεργασία εικόνας, η επεξεργασία σήματος στις τηλεπικοινωνίες, ο έλεγχος συστημάτων και η επεξεργασία σεισμικών δεδομένων.

Τέλος, η ψηφιακή επεξεργασία σήματος έχει και μειονεκτήματα. Ένα από αυτά είναι και οι περιορισμοί στην ταχύτητα λειτουργίας των μετατροπέων αναλογικού σήματος σε ψηφιακό, καθώς και στους ίδιους τους ψηφιακούς επεξεργαστές σήματος.

3.1.5 Τεχνικές Μηχανικής Μάθησης

Η δημιουργία μοντέλων ή προτύπων από ένα σύνολο δεδομένων, από ένα υπολογιστικό σύστημα, ονομάζεται μηχανική μάθηση[10] (machine learning). Έχουν αναπτυχθεί πολλές τεχνικές μηχανικής μάθησης που χρησιμοποιούνται ανάλογα με τη φύση του προβλήματος και εμπίπτουν σε ένα από τα παρακάτω δυο είδη:

- Μάθηση με επίβλεψη (supervised learning) ή μάθηση με παραδείγματα (learning from examples).
- Μάθηση χωρίς επίβλεψη (unsupervised learning) ή μάθηση από παρατήρηση (learning from observation).

Στη **μάθηση με επίβλεψη** το σύστημα εκπαιδεύεται από ένα δάσκαλο και καλείται να «μάθει» από ένα σύνολο δεδομένων, που αντιπροσωπεύουν χαρακτηριστικά του μοντέλου.

Στη **μάθηση χωρίς επίβλεψη** το σύστημα ομαδοποιεί για κάποιο σύνολο εισόδων χωρίς να γνωρίζει επιθυμητές εξόδους για το σύνολο εκπαίδευσης.

3.1.5.1 Μάθηση με Επίβλεψη

1. Μάθηση Παραμέτρων (Concept Learning)

Η μάθηση παραμέτρων [11] είναι τυπικό παράδειγμα επαγωγικής μάθησης κατά την οποία, το σύστημα καλείται να μάθει με παραδείγματα που ανήκουν (θετικά παραδείγματα) ή δεν ανήκουν (αρνητικά παραδείγματα) στη συγκεκριμένη παράμετρο. Ο πιο γνωστός αλγόριθμος μάθησης παραμέτρων είναι ο αλγόριθμος απαλοιφής υποψηφίων.

Ο Αλγόριθμος Απαλοιφής Υποψηφίων (Candidate Elimination Algorithm)

Περιορίζει το χώρο αναζήτησης επιτελώντας γενικεύσεις και εξειδικεύσεις σε κάποιες αρχικές υποθέσεις (παραμέτρους) με βάση τα δεδομένα εκπαίδευσης. Διατηρεί δύο σύνολα, G και S , που από κοινού περιγράφουν όλο το χώρο αναζήτησης και ορίζονται ως εξής:

G : το σύνολο των πιο γενικών (maximally general) υποψηφίων υποθέσεων (δηλαδή παραμέτρων).

S : το σύνολο των πιο εξειδικευμένων (maximally specific) υποψηφίων υποθέσεων.

2. Δένδρα Ταξινόμησης/Απόφασης

Δενδροειδής δομή που με γραφικό τρόπο περιγράφει τα δεδομένα. Κάθε κόμβος ορίζει μια συνθήκη ελέγχου της τιμής κάποιου χαρακτηριστικού των περιπτώσεων. Κάθε κλαδί που φεύγει από ένα κόμβο αντιστοιχεί σε μια διαφορετική διακριτή τιμή του χαρακτηριστικού που σχετίζεται με τον κόμβο. Στα κλαδιά φύλλα έχουμε το τι συνέβη.

Αλγόριθμος ID3

Είναι ο πιο γνωστός αλγόριθμος μάθησης δένδρων ταξινόμησης και είναι αναδρομικός. Ο ID3[12] κατασκευάζει το δένδρο άπληστα (greedy) από πάνω προς τα κάτω επιλέγοντας αρχικά το πιο κατάλληλο χαρακτηριστικό για έλεγχο στη ρίζα. Στη συνέχεια, για κάθε δυνατή τιμή του χαρακτηριστικού δημιουργούνται οι αντίστοιχοι απόγονοι της ρίζας και τα δεδομένα μοιράζονται στους νέους κόμβους ανάλογα με την τιμή που έχουν για το χαρακτηριστικό που ελέγχεται στη ρίζα. Η όλη διαδικασία επαναλαμβάνεται για κάθε νέο κόμβο. Η διαδικασία τερματίζει όταν οι κόμβοι γίνουν τερματικοί (ή φύλλα).

3. Μάθηση κατά Περίπτωση (Instance-based Learning)

Τα δεδομένα εκπαίδευσης διατηρούνται αμετάβλητα σε αντίθεση με τις άλλες μεθόδους μηχανικής μάθησης οι οποίες κωδικοποιούν τα παραδείγματα εκπαίδευσης σε μια συμπαγή περιγραφή. Χαρακτηριστικότερος είναι ο αλγόριθμος των k-κοντινότερων γειτόνων (k-Nearest Neighbors). Κάνει την παραδοχή ότι τα διάφορα παραδείγματα μπορεί να αναπαρασταθούν ως σημεία σε κάποιον Ευκλείδειο χώρο R^n με n διάσταση, όπου n ο αριθμός των χαρακτηριστικών (ανεξάρτητων μεταβλητών). Οι κοντινότεροι γείτονες μιας περίπτωσης υπολογίζονται με βάση την Ευκλείδεια απόστασή τους.

4. Μάθηση κατά Bayes

Στη μάθηση κατά Bayes [13] (Bayesian learning) κάθε παράδειγμα εκπαίδευσης μπορεί σταδιακά να μειώσει ή να αυξήσει την πιθανότητα να είναι σωστή μια υπόθεση.

Μια πρακτική δυσκολία στην εφαρμογή της μάθησης κατά Bayes είναι η απαίτηση για τη γνώση πολλών τιμών πιθανοτήτων. Υπάρχει ακόμη μια απλουστευμένη εκδοχή της μάθησης κατά Bayes, ο απλός ταξινομητής Bayes, στον οποίο γίνεται η παραδοχή ότι τα χαρακτηριστικά είναι ανεξάρτητα μεταξύ τους.

5.Νευρωνικά Δίκτυα (Neural Networks)

Τα Νευρωνικά Δίκτυα[14] παρέχουν ένα εύκολο τρόπο για την εκμάθηση αριθμητικών και διανυσματικών συναρτήσεων ορισμένων σε συνεχή ή διακριτά μεγέθη. Έχουν το μεγάλο πλεονέκτημα της ανοχής που παρουσιάζουν σε δεδομένα εκπαίδευσης σε θόρυβο, δηλαδή δεδομένα που περιστασιακά έχουν λανθασμένες τιμές (π.χ. λάθη καταχώρησης). Άλλο σημαντικό πλεονέκτημα είναι ότι έχουν την ικανότητα να γενικεύουν. Αδυνατούν όμως να εξηγήσουν ποιοτικά τη γνώση που μοντελοποιούν.

3.1.5.2 Μάθηση χωρίς Επίβλεψη

Κανόνες Συσχέτισης

Η ανακάλυψη ή εξόρυξη κανόνων συσχέτισης (association rule mining) εμφανίστηκε αρκετά αργότερα από τη μηχανική μάθηση και έχει περισσότερες επιρροές από την ερευνητική περιοχή των βάσεων δεδομένων.

Αλγόριθμοι Εύρεσης Κανόνων Συσχέτισης

Για την ανακάλυψη κανόνων συσχέτισης χρησιμοποιείται η ιδιότητα της μονοτονίας (monotonicity property) ή αλλιώς ιδιότητα Apriori [15]. Σε έναν αλγόριθμο εύρεσης κανόνων συσχέτισης μας ενδιαφέρει κυρίως ο αριθμός των περασμάτων στα δεδομένα που απαιτείται κατά την εκτέλεσή του.

Αλγόριθμος Apriori

Περιλαμβάνει δυο βασικά βήματα, τη δημιουργία των συχνών συνόλων αντικειμένων και τη δημιουργία των κανόνων συσχέτισης. Για τη δημιουργία των κανόνων συσχέτισης ελέγχεται η εμπιστοσύνη (confidence) όλων των πιθανών κανόνων που προκύπτουν από τα μέγιστα συχνά σύνολα αντικειμένων και στο τέλος μένουν εκείνοι των οποίων η εμπιστοσύνη ξεπερνά το όριο που τέθηκε από το χρήστη.

3.1.5.3 Άλλα Είδη Μάθησης

Εκτός από τις μεθόδους που παρουσιάστηκαν, υπάρχουν και άλλες προσεγγίσεις στο πρόβλημα της μηχανικής μάθησης. Δύο από αυτές, είναι οι γενετικοί αλγόριθμοι και η ενισχυτική μάθηση.

Γενετικοί Αλγόριθμοι (Genetic Algorithms)

Είναι μια μέθοδος μάθησης που βασίζεται στην προσομοίωση του φυσικού φαινομένου της εξέλιξης (evolution). Οι υποθέσεις συνήθως αναπαριστώνται από ακολουθίες bit (bit-strings). Η αναζήτηση της κατάλληλης υπόθεσης ξεκινάει τυχαία με έναν πληθυσμό (μια συλλογή) αρχικών υποθέσεων, τα μέλη του οποίου παράγουν τη νέα «γενιά» μέσω διαδικασιών αναπαραγωγής αντίστοιχων των βιολογικών, όπως: επιλογή (selection), διασταύρωση (crossover) και τυχαία μετάλλαξη (random mutation).

Ενισχυτική Μάθηση (Reinforcement Learning)

Η ενισχυτική μάθηση [16] είναι μια γενική περιγραφή οικογένειας τεχνικών στις οποίες το σύστημα μάθησης προσπαθεί να μάθει μέσω άμεσης αλληλεπίδρασης με το περιβάλλον. Είναι εμπνευσμένη από τα αντίστοιχα ανάλογα της μάθησης με επιβράβευση και τιμωρία που συναντώνται στα έμβια όντα μέχρι να επέλθει η μάθηση. Το σύστημα δεν καθοδηγείται από κάποιον δάσκαλο για το ποια ενέργεια θα πρέπει να ακολουθήσει αλλά πρέπει να ανακαλύψει μόνο του ποιες ενέργειες είναι αυτές που θα του αποφέρουν το μεγαλύτερο κέρδος.

Κεφάλαιο 4

Ανάλυση απαιτήσεων και προδιαγραφές

4.1 Συναρτήσεις Matlab

15

4.1 Συναρτήσεις Matlab

Σε αυτό το υποκεφάλαιο θα εξηγήσουμε διάφορες έτοιμες συναρτήσεις που είναι υλοποιημένες στη Matlab, τις οποίες χρησιμοποίησα για την υλοποίηση της παρούσας Ατομικής Διπλωματικής Εργασίας.

Wavread

Φορτώνει ένα αρχείο τύπου «wave». Δέχεται ως είσοδο το όνομα του αρχείου και επιστρέφει στην έξοδο τα «samples» (δείγματα) των δεδομένων.

Clear

Αφαιρεί όλες τις μεταβλητές από το χώρο εργασίας και τις αποδεσμεύει από την μνήμη.

Cle

Καθαρίζει το χώρο εργασίας.

If

Είναι μια δομή διακλάδωσης η οποία εκτελεί τις ενέργειες που υπάρχουν στο μπλοκ της αν ισχύει μια συνθήκη.

For

Είναι μια δομή επανάληψης η οποία εκτελεί τις ενέργειες που υπάρχουν στο μπλοκ της σε μια ή περισσότερες επαναλήψεις μέχρι να τερματιστεί η συνθήκη της. Στη συνθήκη

της περιλαμβάνεται η αρχικοποίηση ενός μετρητή και η αύξηση του μετρητή γίνεται αυτόματα (σε κάθε επανάληψη αυξάνεται κατά ένα), εκτός αν δηλωθεί διαφορετικά.

While

Είναι μια δομή επανάληψης που εκτελεί την ίδια λειτουργία με την εντολή «for», με τη διαφορά όμως ότι η αρχικοποίηση του μετρητή που περιλαμβάνεται στη συνθήκη της πρέπει να γίνει πριν την εντολή «while» και η αύξηση του μετρητή της γράφεται συνήθως στο τέλος του μπλοκ των εντολών της.

Load

Φορτώνει όλες τις μεταβλητές από ένα .mat αρχείο και τις αποθηκεύει σε ένα πίνακα. Παίρνει σαν παράμετρο το όνομα αρχείου.

Importdata

Φορτώνει δεδομένα από ένα αρχείο στον χώρο εργασίας. Παίρνει σαν παράμετρο όνομα αρχείου.

Cd

Μας μεταφέρει σ' ένα νέο φάκελο.

Plot

Δέχεται ως είσοδο ένα πίνακα ή ένα διάνυσμα. Παρουσιάζει τα δεδομένα του πίνακα ή του διανύσματος στον άξονα των ψ ως προς τον άξονα των χ που είναι η αρίθμηση των δεδομένων αυτών.

Subplot

Χωρίζει την παρούσα απεικόνιση σε ορθογώνια παράθυρα που αριθμούνται ως προς τις γραμμές. Κάθε παράθυρο περιέχει ένα αντικείμενο με άξονες που μπορούμε να χειριστούμε χρησιμοποιώντας τις ιδιότητες του.

Ylabel

Βάζουμε όνομα στον άξονα των Ψ . Παίρνει σαν παράμετρο το όνομα που θέλουμε να εμφανίζεται στον άξονα Ψ .

Xlabel

Βάζουμε όνομα στον άξονα των X. Παίρνει σαν παράμετρο το όνομα που θέλουμε να εμφανίζεται στον άξονα X.

Hold on

Διατηρεί το τρέχον γράφημα και προσθέτει άλλο ένα γράφημα σε αυτό. Η MATLAB προσαρμόζει τα όρια στους άξονες και σημειώνει τις ετικέτες όπως είναι απαραίτητο για να εμφανιστεί το πλήρες φάσμα του γραφήματος που προστίθεται.

Hold off

Επαναφέρει την κατάσταση «hold» στην προεπιλεγμένη συμπεριφορά, στην οποία η MATLAB καθαρίζει το υπάρχον γράφημα και επαναφέρει τις ιδιότητες των αξόνων στις προεπιλογές τους πριν την σχεδίαση νέων γραφημάτων.

Length

Εάν έχουμε μονοδιάστατο πίνακα βρίσκει το μέγεθος του, αλλιώς αν είναι πολυδιάστατος βρίσκει το μέγεθος της μεγαλύτερης διάστασής του.

Zeros

Δημιουργεί έναν πίνακα με δεδομένα σε κάθε θέση τον αριθμό 0. Παίρνει σαν παράμετρο τη διάσταση που θέλουμε να έχει ο πίνακας.

Min

Επιστρέφει το μικρότερο στοιχείο από έναν πίνακα. Παίρνει σαν παράμετρο το όνομα πίνακα.

Max

Επιστρέφει το μεγαλύτερο στοιχείο από έναν πίνακα. Παίρνει σαν παράμετρο το όνομα πίνακα.

Abs

Επιστρέφει έναν πίνακα ο οποίος περιέχει στοιχεία σε απόλυτη τιμή. Παίρνει σαν παράμετρο το όνομα του πίνακα.

Strcmp

Συγκρίνει δύο συμβολοσειρές για ισότητα. Οι συμβολοσειρές θεωρείται ότι είναι ίσες, αν το μέγεθος και το περιεχόμενο της κάθε μιας είναι το ίδιο. Η συνάρτηση επιστρέφει τον αριθμό 1 για ισότητα των συμβολοσειρών ή τον αριθμό 0 για την ανισότητα των συμβολοσειρών.

Find

Δέχεται ως είσοδο ένα πίνακα X. Εντοπίζει όλα τα μη μηδενικά στοιχεία του πίνακα X και επιστρέφει τους γραμμικούς δείκτες των στοιχείων αυτών σε ένα μονοδιάστατο πίνακα. Εάν το X είναι ένα διάνυσμα γραμμής, τότε επιστρέφει ένα διάνυσμα γραμμής, διαφορετικά ένα διάνυσμα στήλης. Αν ο X δεν περιέχει μη μηδενικά στοιχεία ή είναι ένας άδειος πίνακας, τότε επιστρέφει έναν άδειο πίνακα.

Ceil

Στρογγυλοποιεί τα στοιχεία ενός πίνακα A που δέχεται ως είσοδο στο πλησιέστερο ακέραιο αριθμό μεγαλύτερο ή ίσο με τον A.

Repmat

Αναπαράγει ένα πίνακα A που δέχεται ως είσοδο σε ένα πίνακα B που αποτελείται από $m \times n$ αντίτυπα του A.

Im2col

Δέχεται ως είσοδο μια εικόνα και αναδιατάσσει μπλοκ της εικόνας σε διανύσματα στηλών.

Exp

Επιστρέφει την εκθετική τιμή της μεταβλητής που δέχεται ως είσοδο. Αν δεχθεί ως είσοδο πίνακα επιστρέφει την εκθετική τιμή κάθε στοιχείου του πίνακα σε ένα νέο πίνακα.

Geomean

Υπολογίζει την γεωμετρική μέση τιμή ενός δείγματος. Για μονοδιάστατους πίνακες ως είσοδο, επιστρέφει τον γεωμετρικό μέσο των στοιχείων του. Για δυσδιάστατους πίνακες ως είσοδο, επιστρέφει ένα διάνυσμα γραμμής που περιέχει τα γεωμετρικά μέσα της κάθε στήλης του πολυδιάστατου πίνακα.

Sum

Δέχεται ως είσοδο ένα πίνακα. Αν είναι μονοδιάστατος επιστρέφει το άθροισμα των στοιχείων του, αν είναι δυσδιάστατος επιστρέφει ένα διάνυσμα γραμμής που περιέχει το άθροισμα των στηλών του.

Isempty

Δέχεται ως είσοδο ένα πίνακα και επιστρέφει τον αριθμό 1 αν είναι κενός, διαφορετικά τον αριθμό 0.

Xlsread

Διαβάζει τα δεδομένα από το πρώτο φύλλο εργασίας από ένα αρχείο Microsoft Excel με είσοδο το όνομα αρχείου. Επιστρέφει τα αριθμητικά δεδομένα σε ένα πίνακα, τα πεδία κειμένου σε ένα άλλο πίνακα και τα μη επεξεργασμένα δεδομένα (αριθμούς και κείμενα) σε ένα τρίτο πίνακα.

Str2double

Μετατρέπει τη συμβολοσειρά που δέχεται ως είσοδο σε μια αναπαράσταση πραγματικού αριθμού. Η συμβολοσειρά θα πρέπει να είναι σε μια αναπαράσταση χαρακτήρα τύπου ASCII που αντιστοιχεί σε ένα πραγματικό αριθμό.

Tic

Αρχίζει ένα χρονόμετρο για να μετρήσει μια επίδοση.

Toc

Σταματά το χρονόμετρο που ξεκίνησε η συνάρτηση tic και παρουσιάζει το χρόνο που πέρασε σε δευτερόλεπτα.

Mkdir

Δημιουργεί ένα καινούριο φάκελο. Το όνομα του φακέλου το δέχεται ως είσοδο και μπορεί να είναι μια απόλυτη ή μια σχετική διαδρομή.

Get

Επιστρέφει πεδία των αντικειμένων του MIRtoolbox όπως: διάρκεια ηχητικού σήματος, όνομα αρχείων, τύπος του αρχείου κτλ.

Connect short pauses

Ενώνει τα κενά όπου η διαφορά τους δεν ξεπερνά το κατώφλι που έχουμε θέσει. Παίρνει σαν είσοδο τρεις παραμέτρους.

Κεφάλαιο 5

Σχεδιασμός Συστήματος, Υλοποίηση

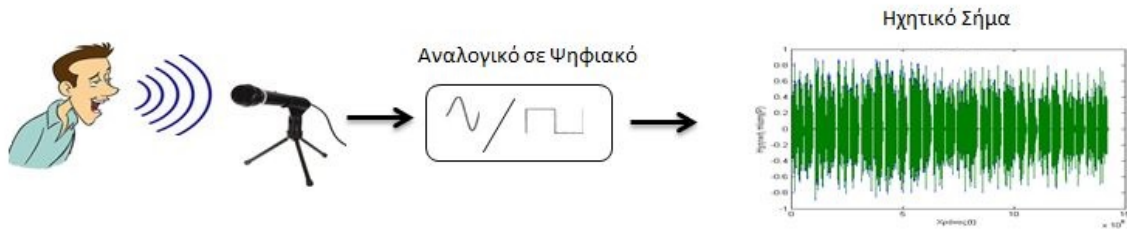
5.1 Μεθοδολογία	21
5.2 Ο αλγόριθμος YIN	23
5.3 Προ-επεξεργασία σήματος	24
5.4 Σχεδιασμός φίλτρων COSFIRE	27
5.5 Δυναμική Στρέβλωση Χρόνου (Dynamic Time Warping - DTW)	30
5.6 Επίτευξη ανοχής συχνότητας	30

5.1 Μεθοδολογία

Ψηφιακή αναπαράσταση του ήχου:

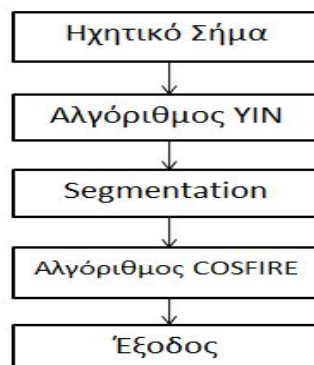
Ο ήχος δημιουργείται όταν ένα υλικό ταλαντώνεται. Με την ταλάντωση του υλικού, τα μόρια του αέρα που βρίσκονται στα τοιχώματα του υλικού αναγκάζονται να συμπυκνώνονται και να αραιώνονται δημιουργώντας έτσι σημεία υψηλής και χαμηλής ηχητικής πίεσης. Στο σχήμα 5.1.1, παρουσιάζεται ένας άνθρωπος ο οποίος τραγουδά. Με μπλε παράλληλες γραμμές αναπαρίσταται το ηχητικό κύμα. Η συνηθέστερη διαδικασία λήψης, ψηφιοποίησης και αποθήκευσης του ήχου γίνεται με την βοήθεια ενός μικροφώνου, ενός μετασχηματιστή του αναλογικού σήματος σε ψηφιακό και έναν υπολογιστή.

Το μικρόφωνο μετατρέπει την ηχητική πίεση σε αναλογικό ρεύμα όπου με τη βοήθεια ενός καλωδίου μεταφέρεται στον μετασχηματιστή (στην περίπτωση μας είναι κάρτα ήχου) για να ψηφιοποιηθεί. Μετά την ψηφιοποίηση ενός τραγουδιού, η ψηφιακή αναπαράσταση της ηχητικής πίεσης παρουσιάζεται στο ίδιο σχήμα μετά την έξοδο του μετασχηματιστή. Σε αυτό εδώ το σχήμα έχουμε το αρχικό ψηφιακό ηχητικό σήμα ενός τραγουδιού όπου ο άξονας Ψ αντιπροσωπεύει την ηχητική πίεση (P) σε συνάρτηση με τον χρόνο (t) όπου είναι ο άξονας X .



Σχήμα 5.1.1. Ψηφιακή αναπαράσταση ήχου

Στο σχήμα 5.1.2 βλέπουμε την διαδικασία της μεθοδολογίας που ακολουθήθηκε. Το ηχητικό σήμα δίδεται στον αλγόριθμο YIN [17] όπου στην έξοδο δίδονται οι θεμελιώδεις συχνότητες του σήματος στον χρόνο. Στη συνέχεια, γίνεται κάποια κατάτμηση του σήματος σε πιθανά όρια που καθορίζουν την αρχή και το τέλος κάθε στροφής που θα συγκρίνεται. Το σήμα της συχνότητας και τα σημεία της κατάτμησης, καθώς και κάποιες παράμετροι, δίδονται στην είσοδο του αλγόριθμου COSFIRE. Στην έξοδο του αλγόριθμου παίρνουμε τις πληροφορίες που σχετίζονται με την επαναληψιμότητα των στροφών.



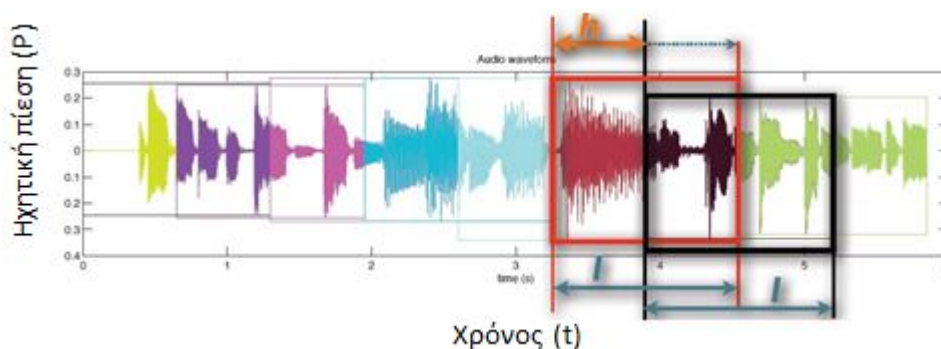
Σχήμα 5.1.2. Διαδικασία εξαγωγής όμοιων φράσεων

Για να μπορέσουμε να επεξεργαστούμε τις φράσεις κάποιου τραγουδιού χρειαζόμαστε την πληροφορία της θεμελιώδους συχνότητας σε συνάρτηση με τον χρόνο. Ο λόγος που χρειάζεται να μετατρέψουμε το ηχητικό σήμα της πίεσης σε σήμα της συχνότητας είναι γιατί η συχνότητα αντιπροσωπεύει την μελωδία. Ο αλγόριθμος YIN δέχεται στην είσοδο του το ηχητικό σήμα πίεση-χρόνος, καθώς και διάφορες παραμέτρους. Στην έξοδο του δίνει τις τιμές της θεμελιώδους συχνότητας.

5.2 Ο αλγόριθμος YIN

Για να μπορέσουμε να έχουμε μια αναπαράσταση της θεμελιώδους συχνότητας ως προς τον χρόνο, θα χρειαστεί να κατατμηθεί το ηχητικό σήμα σε πολλά πλαίσια συγκεκριμένου μεγέθους. Με την έννοια πλαίσιο εννοούμε την ανάλυση όλου του χρονικού σήματος όπου πραγματοποιείται σε ένα πλαίσιο που κινείται χρονικά κατά μήκος του χρονικού σήματος. Μπορεί να έχει μέγεθος 64, 128, 256, 512, 1024, 2048 κλπ. θέσεις πίνακα.

Για κάθε πλαίσιο, ο αλγόριθμος YIN εφαρμόζει την τεχνική αυτό-συσχέτισης και με κάποια περεταίρω διαδικασία εξάγει μία συχνότητα, όπου είναι η θεμελιώδης συχνότητα του σήματος. Επίσης χρησιμοποιείται και το hop (h) το οποίο είναι η απόσταση 2 διαδοχικών πλαισίων επικαλύπτοντας με αυτό τον τρόπο τα διαδοχικά πλαίσια, όπως φαίνεται στο σχήμα 5.2.1.

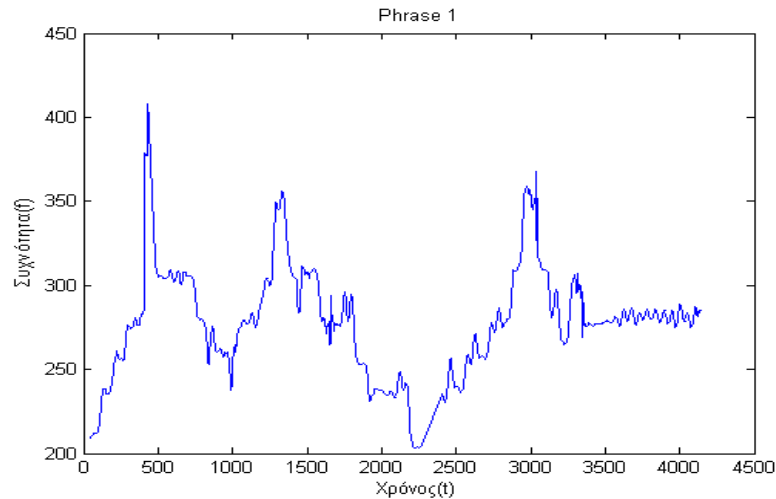


Σχήμα 5.2.1. Στο σχήμα βλέπουμε την διαδικασία όπου περνάμε ως είσοδο στον αλγόριθμο YIN το audio σήμα έτσι όπως έχουν κατατμιστεί τα διάφορα πλαίσια ανάλογα με το size που θέσαμε

Στην έξοδο του αλγόριθμου θα έχουμε πολλές τιμές συχνότητας όσες δηλαδή θα είναι και τα πλαίσια που θα δημιουργηθούν ανάλογα με το μέγεθος του τραγουδιού και το μέγεθος του πλαισίου που έχουμε θέσει.

Συνήθως το ηχητικό σήμα είναι κωδικοποιημένο με 44100 δείγματα το δευτερόλεπτο. Αν θέσουμε ένα πλαίσιο 128 δειγμάτων, τότε θα έχουμε $44100/128$ πλαίσια το δευτερόλεπτο με αυτό τον τρόπο μικραίνει ο όγκος των δεδομένων.

Στην συνέχεια καθώς έχει τελειώσει ο αλγόριθμος και έχουμε πάρει τις συχνότητες F0 και καθώς είναι φυλαγμένες σε ένα πίνακα τότε η γραφική που θα έχουμε φαίνεται στο σχήμα 5.2.2. Ο άξονας των X θα είναι χρόνος (t) και ο άξονας τον Ψ η συχνότητα (f).

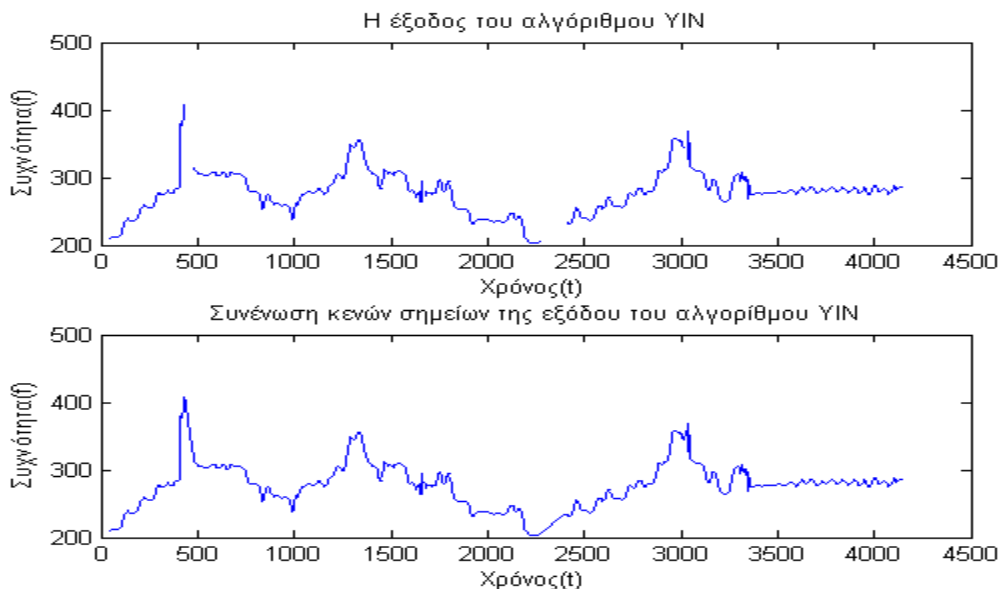


Σχήμα 5.2.2. Γραφική παράσταση εξόδου YIN της συχνότητας των στροφών συναρτήσει του χρόνου

5.3 Προ-επεξεργασία σήματος

Όπως διακρίνουμε από την γραφική σχήμα 5.3.1(a) , μεταξύ των στροφών υπάρχει κάποιο χρονικό διάστημα όπου δεν υπάρχουν τιμές και αυτό οφείλεται στις ανάσες που παίρνει ο τραγουδιστής ή στο (octave error).

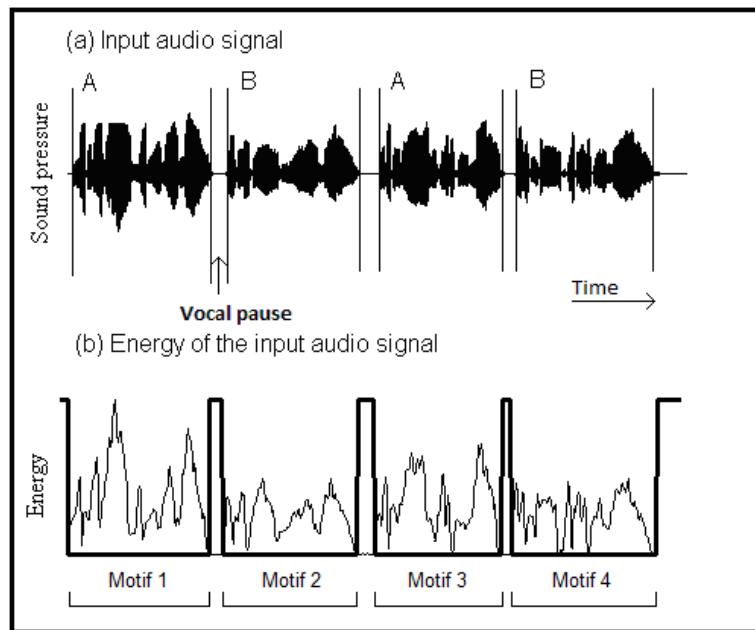
Συνήθως οι στροφές αρχίζουν και τελειώνουν όπου υπάρχουν κενά φωνής. Έτσι εμείς για να κάνουμε ένα αυτοματοποιημένο σύστημα για αναγνώριση των στροφών αποφασίσαμε να χρησιμοποιήσουμε ως κριτήριο την παρουσία αυτών των κενών. Με την βοήθεια μιας συνάρτησης θα περάσουμε ως είσοδο την έξοδο του αλγόριθμου YIN και έτσι θα ενώσει τα σημεία που δεν ενώνονται και η διαφορά των σημείων αυτών δεν ξεπερνά το κατώφλι που έχουμε θέσει, όπως φαίνεται στο σχήμα 5.3.1(b).



Σχήμα 5.3.1. (a) Η έξοδος του αλγορίθμου YIN, (b) Συνένωση κενών σημείων της εξόδου του αλγορίθμου YIN

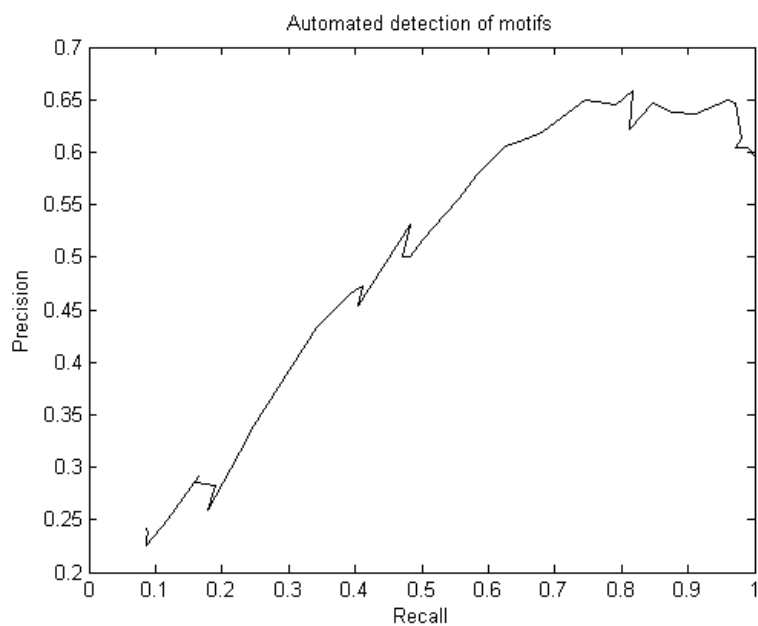
Στο [19] από τη βιβλιογραφία, οι συγγραφείς προτείνουν μια μέθοδο για την ανεύρεση μικρογραφίας ήχου της λαϊκής μουσικής. Μια μικρογραφία ήχου θεωρείται ως το πιο αντιπροσωπευτικό κομμάτι της μουσικής. Για τη λαϊκή μουσική, το ρεφρέν θεωρείται ότι είναι μια ιδανική μικρογραφία ήχου. Η μέθοδός τους περιλαμβάνει μια κατάτμηση ήχου βασισμένη σε πλαίσια, όπου κάθε πλαίσιο ήχου περιγράφεται με ένα σύνολο χαρακτηριστικών. Στη συνέχεια, υπολογίζονται οι πίνακες ομοιότητας των εξαγόμενων χαρακτηριστικών και χρησιμοποιούνται για τον εντοπισμό φράσεων που επαναλαμβάνονται σε ολόκληρο το σήμα ήχου.

Πιο συγκεκριμένα, το σήμα ήχου χωρίζεται σε μία ακολουθία επικαλυπτόμενων πλαισίων με 1024 bins για πλάτος (24 ms) και 128 bins για την επικάλυψη (3 ms). Για κάθε τέτοιο πλαίσιο ήχου εξάγουμε την ενέργεια του σήματος. Στο σχήμα 5.3.2. (b), η ενέργεια του ηχητικού σήματος παριστάνεται γραφικά με μια λεπτή συνεχή γραμμή. Η κατηγοριοποίηση σε κομμάτια όπου υπάρχει τραγούδι ή σε κομμάτια όπου υπάρχει σιωπή γίνεται με ένα κατώφλι στις τιμές της ενέργειας. Εμείς μετατρέπουμε τις τιμές του μονοδιάστατου πίνακα της ενέργειας σε δυαδικές και ορίζουμε μια τιμή ίση με 1 εάν ένα πλαίσιο ήχου έχει τιμή ενέργειας χαμηλότερη από το κατώφλι. Στην περίπτωση αυτή, το πλαίσιο έχει κατηγοριοποιηθεί ως κομμάτι όπου υπάρχει σιωπή. Διαφορετικά ορίζουμε μια τιμή ίση με 0. Ο δυαδικός μονοδιάστατος πίνακας της ενέργειας φαίνεται στο σχήμα 5.3.2. (b) με μια παχιά συνεχή γραμμή.



Σχήμα 5.3.2. (a) Γραφική παράσταση των τιμών του ηχητικού σήματος έναντι του χρόνου για X δευτερόλεπτα. Οι κάθετες γραμμές οριοθετούν τις τέσσερις φράσεις. (b) Η ενέργεια του ηχητικού σήματος. Το δυαδικό διάγραμμα δείχνει τη θετική κατηγοριοποίηση των τμημάτων που περιέχουν σιωπή.

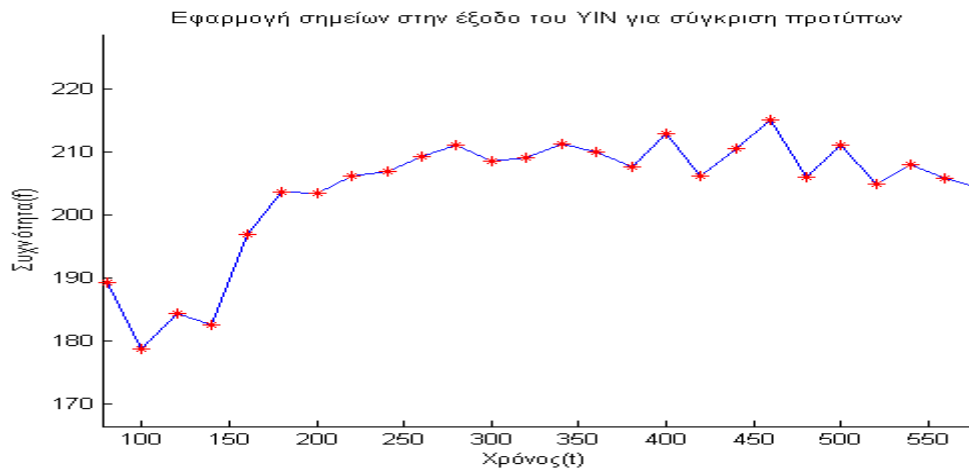
Η διάρκεια των φωνητικών παύσεων που εμφανίζονται μέσα σε μια φράση είναι συνήθως μικρότερη από τις φωνητικές παύσεις μεταξύ διαδοχικών φράσεων. Η χρήση ενός κατώφλιου για την ανίχνευση πιθανής θέσης ορίων μιας φράσης οδηγεί σε λανθασμένες κατηγοριοποιήσεις όταν μια φωνητική παύση με διάρκεια μεγαλύτερη από το κατώφλι εμφανίζεται μέσα σε μια φράση. Στο σχήμα 5.3.3. παρουσιάζουμε τα αποτελέσματα της ανίχνευσης των φράσεων με βάση τη διάρκεια των φωνητικών παύσεων.



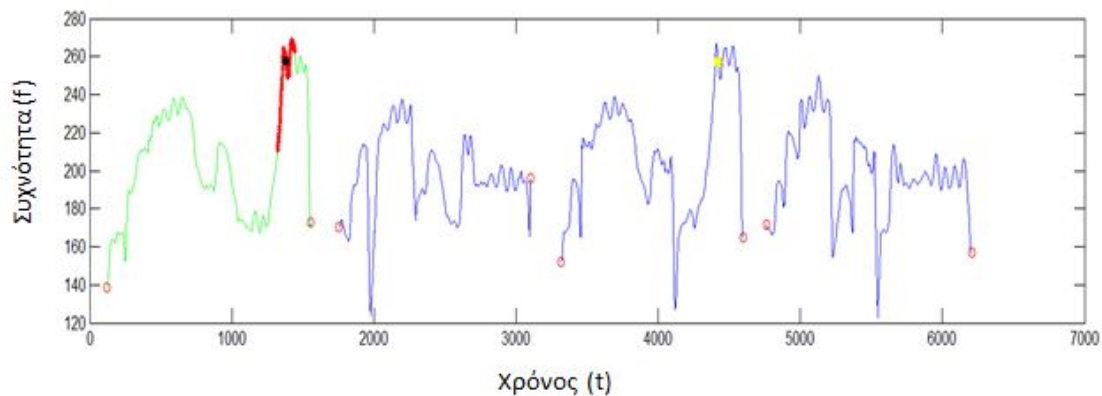
Σχήμα 5.3.3. Η ακρίβεια P και η ανάκληση R σχετικά με τον προσδιορισμό των θέσεων έναρξης και λήξης των φράσεων. Ο συνολικός αριθμός των φράσεων είναι 123.

5.4 Σχεδιασμός φίλτρων COSFIRE

Για να μπορέσουμε να συγκρίνουμε τις φράσεις θα τις χωρίσουμε ανά 20 σημεία, όπως φαίνεται στο σχήμα 5.4.1. Για κάθε σημείο, παίρνουμε την τιμή συχνότητας που αναφέρεται στο συγκεκριμένο σημείο, όπου θα είναι το κεντρικό σημείο του φίλτρου COSFIRE. Από το κεντρικό σημείο καθορίζουμε 60 σημεία αριστερά και 60 δεξιά. Αυτό το σήμα με τα 121 σημεία το ονομάζουμε πρωτότυπο όπως φαίνεται στο σχήμα 5.4.2 με κόκκινη γραμμή. Στο ίδιο σχήμα, με μαύρο φαίνεται το κεντρικό σημείο του φίλτρου.



Σχήμα 5.4.1. Εφαρμογή κεντρικών σημείων στο σήμα της συχνότητας.



Σχήμα 5.4.2. Στο σχήμα με κόκκινο στρογγυλό φαίνεται η αρχή και το τέλος κάθε στροφής που ανιχνεύτηκε.

Ακολουθώντας, παίρνουμε κάθε φορά ένα πρωτότυπο από την φράση εκπαίδευσης για να μπορούμε να συγκρίνουμε με το ανάλογο πρωτότυπο από μια άλλη φράση επαλήθευσης.

Ο αλγόριθμος COSFIRE για την σύγκριση κάθε πρωτοτύπου με κάποιο άλλο χρησιμοποιεί την εξίσωση:

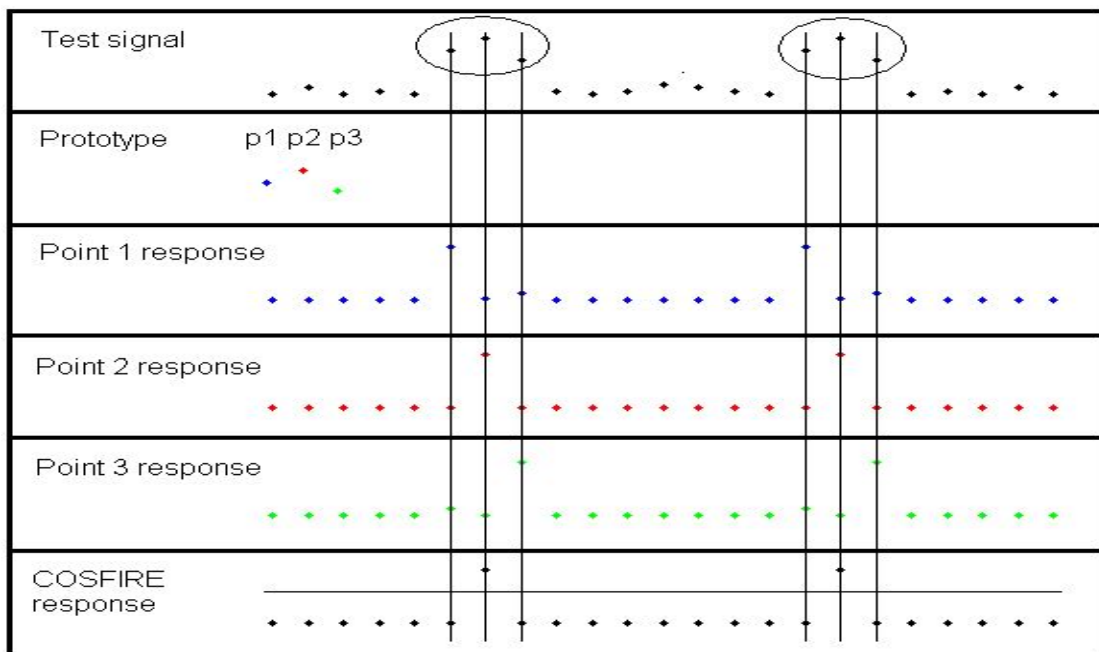
$$d = e^{-\frac{(\text{prototype}(f) - \text{test}(f))^2}{2\sigma^2}} \quad (1)$$

η οποία εάν η διαφορά των 2 σημείων είναι 0 τότε ξέρουμε ότι e^0 ισούται με 1. Αυτό το βλέπουμε από την απάντηση πιο κάτω του κάθε πρωτοτύπου P1, P2 και P3 (point response).

Για κάθε σημείο του πρωτοτύπου υπολογίζουμε την απόσταση που φαίνεται στην εξίσωση (1), μια μέτρηση απόστασης από μια συνάρτηση Gaussian, των οποίων η

Sigma είναι μια γραμμική συνάρτηση της απόστασης από το κέντρο του φίλτρου.

Ένα παράδειγμα για το πώς λειτουργεί η προτεινόμενη μέθοδος COSFIRE απεικονίζεται στο σχήμα 5.4.3. Ας θεωρήσουμε ένα συνεχή σήμα με θόρυβο που εμφανίζεται στο πρώτο κουτί του ίδιου σχήματος. Αυτό το σήμα περιέχει δύο περιπτώσεις που μας συμφέρει να τα εντοπίσει αυτόματα. Στο δεύτερο κουτί, διαμορφώσαμε ένα πρωτότυπο χρησιμοποιώντας τρεις τιμές που μας ενδιαφέρουν. Το κέντρο του φίλτρου βρίσκεται στη μέση του πρωτοτύπου και εμφανίζεται με κόκκινο χρώμα. Οι αποκρίσεις των Gaussian τιμών απόστασης για τα τρία σημεία φαίνονται στα επόμενα τρία κουτιά. Οι τρεις αποκρίσεις των Gaussian τιμών απόστασης μετατοπίζονται προς το κέντρο του φίλτρου COSFIRE και ο γεωμετρικός μέσος φαίνεται στο τελευταίο κουτί. Η απόκριση του φίλτρου COSFIRE έχει τιμή 0,98 στο κέντρο των δύο σημείων που μας ενδιαφέρουν και τιμή 0 οπουδήποτε αλλού. Η απόκριση τελικά γίνεται δυαδική με ένα κατώφλι που φαίνεται με μια μαύρη συνεχόμενη γραμμή στο κουτί απόκρισης COSFIRE.



Σχήμα 5.4.3. Ρύθμιση και απόκριση ενός φίλτρου COSFIRE. Στο πρώτο κουτί φαίνεται το σήμα εισόδου. Τα δύο σημεία που μας ενδιαφέρουν φαίνονται στο ίδιο κουτί με τους κύκλους. Το πρωτότυπο που χρησιμοποιείται για να εκπαιδεύσει το φίλτρο φαίνεται στο δεύτερο κουτί. Τα επόμενα 3 κουτιά δείχνουν τις αποκρίσεις των τριών σημείων του πρωτοτύπου. Το τελευταίο κουτί δείχνει την απόκριση COSFIRE.

5.5 Δυναμική Στρέβλωση Χρόνου (Dynamic Time Warping - DTW)

Εφαρμόσαμε τον αλγόριθμο DTW [18] στα δεδομένα και υπολογίστηκε η συνάρτηση κόστους για τα ζεύγη των μοτίβων. Η έξοδος της συνάρτησης κόστους χρησιμοποιήθηκε για να ταξινομήσει τα μοτίβα που δοκιμάστηκαν ως "παρόμοια" ή "ανόμοια". Τα αποτελέσματα του DTW όσον αφορά την ακρίβεια και ανάκληση ήταν συγκρίσιμα με τα αποτελέσματα της μεθόδου COSFIRE. Το μεγαλύτερο μειονέκτημα του DTW είναι ο μεγάλος χρόνος εκτέλεσης.

5.6 Επίτευξη ανοχής συχνότητας

Έχουμε εισαγάγει την παράμετρο «άλφα» η οποία είναι υπεύθυνη για την ανοχή που επιτρέπεται στην απόσταση των συχνότητων μεταξύ του πρωτοτύπου και των δοκιμαζόμενων σημάτων. Η «άλφα» παράμετρος είναι μια τιμή στο εύρος $[0 - 1]$. Η τιμή αυτή προστίθεται στην τυπική απόκλιση της Εξ. 1 για κάθε σημείο μακριά από το κέντρο του φίλτρου COSFIRE. Ως εκ τούτου, η συχνότητα στο κέντρο του φίλτρου έχει λιγότερη ανοχή στη συνάρτηση απόστασης σε σχέση με τα τελευταία ακραία σημεία. Βρήκαμε ότι η βέλτιστη τιμή του «άλφα» που μεγιστοποιεί τον αρμονικό μέσο της ακρίβειας και της ανάκλησης ότι είναι το X . Η τυπική απόκλιση της Gaussian συνάρτησης δεν παίζει κανένα σημαντικό ρόλο στα αποτελέσματα. Ως εκ τούτου, χρησιμοποιούμε μια τυπική τιμή 0.5.

Κεφάλαιο 6

Αξιολόγηση Συστήματος – Αποτελέσματα

6.1 Αξιολόγηση Συστήματος – Αποτελέσματα

31

6.1 Αξιολόγηση Συστήματος – Αποτελέσματα

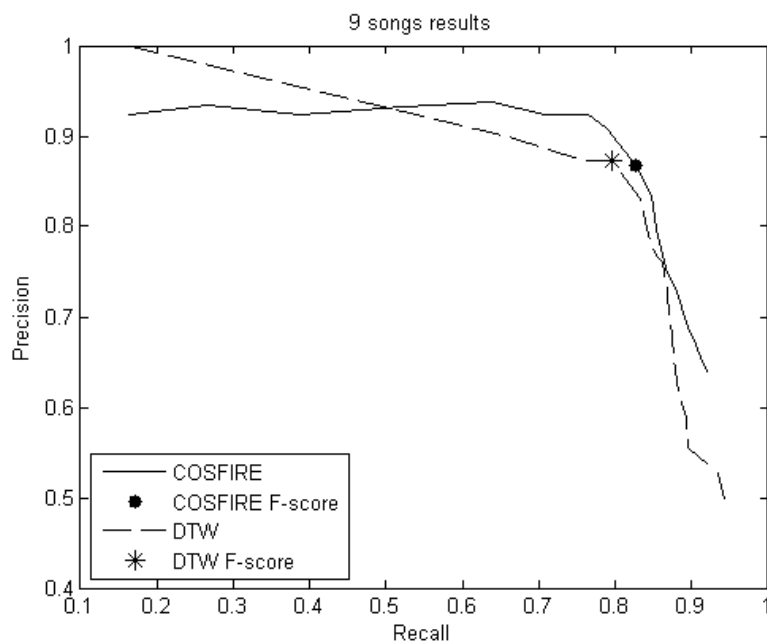
Χρησιμοποιήθηκαν 9 μονοφωνικά λαϊκά τραγούδια για την επικύρωση των μεθόδων COSFIRE και DTW. Ελέγχθηκαν 123 μοτίβα τα οποία εγκρίθηκαν. Στον πίνακα 1 παρουσιάζονται τα δεδομένα και οι ιδιότητες τους. Η βάση δεδομένων περιέχει μονοφωνικά λαϊκά τραγούδια της Κύπρου. Κάθε τραγούδι εκτελείται από διαφορετικό πρόσωπο σε μια διαφορετική χρονική στιγμή και τοποθεσία. Μερικά από τα τραγούδια φαίνεται να έχουν το ίδιο όνομα, αλλά το περιεχόμενο των μελωδιών είναι διαφορετικό.

Τραγούδι	Αριθμός των μοτίβων	Μέσος χρόνος μοτίβων (s)	Τυπική Απόκλιση Διάρκειας Μοτίβων(s)	Διάρκεια τραγουδιού (s)
Gia tin kerinia	18	5.6	0.66	116.6
O pramateftis	23	4.7	0.36	121.8
Afkoritissa 1	8	5.9	2.68	53.1
Nanourisma 1	17	5.2	1.20	98.6
Nanourisma 2	14	4.0	0.31	67.7
Nanourisma 3	15	3.8	0.64	67.6
Nanourisma 4	18	6.6	1.18	129.0
Tis sousas 1	6	9.7	0.26	65.0
Tis sousas 2	4	11.6	0.31	53.3

Πίνακας 6.1. Τα ονόματα των τραγουδιών, η μέση τιμή και η τυπική απόκλιση της διάρκειας των μοτίβων και η συνολική διάρκειά τους.

Παρουσιάζουμε τα αποτελέσματα για τον COSFIRE και τον DTW λαμβάνοντας υπόψη την ακρίβεια (Precision - P) και την ανάκληση (Recall - R). Οι διαφορετικές τιμές της P και της R για τον COSFIRE λήφθηκαν με την αλλαγή της παραμέτρου «άλφα». Χρησιμοποιήσαμε 17 διαφορετικές τιμές της παραμέτρου «άλφα» στο εύρος [0,1 – 0,9]. Η P μειώνεται και η R αυξάνεται με την αύξηση των τιμών του «άλφα». Ο αρμονικός μέσος της ακρίβειας και της ανάκλησης φθάνει σ' ένα μέγιστο σε τιμή της ανάκλησης R 82,7% και τιμή της ακρίβειας P 86,8% για τιμή του «άλφα» = 0,5.

Για τη μέθοδο DTW η κατηγοριοποίηση έγινε με τον καθορισμό ενός κατωφλίου για τη συνάρτηση κόστους της εξόδου του DTW. Χρησιμοποιήσαμε 17 διαφορετικές τιμές του κατωφλίου στην κλίμακα [0-0,8]. Ο αρμονικός μέσος της ακρίβειας και της ανάκλησης φθάνει σ' ένα μέγιστο σε τιμή της ανάκληση R 79,6% και της ακρίβειας P 87,2% για τιμή του κατωφλίου ίση με 0,4. Ο μέγιστος αρμονικός μέσος για τον COSFIRE έχει υψηλότερη αξία σε σύγκριση με εκείνον του DTW με τιμές 0,874 και 0,8325 αντίστοιχα. Από τις τιμές αυτές καταλήγουμε στο συμπέρασμα ότι η μέθοδος COSFIRE αποδίδει καλύτερα από τη μέθοδο DTW.



Σχήμα 6.1. Η γραφική παράσταση της ακρίβειας και της ανάκλησης φαίνεται πιο πάνω όπως λαμβάνεται από τις μεθόδους COSFIRE και DTW. Η ακρίβεια και η ανάκληση για τον COSFIRE εμφανίζονται με μαύρη συνεχή γραμμή και η θέση του μέγιστου αρμονικού μέσου της ακρίβειας και της ανάκλησης απεικονίζεται με μαύρη

κουκίδα. Εμείς επικυρώσαμε τη μέθοδο για διαφορετικές τιμές της παραμέτρου «άλφα». Η ακρίβεια και η ανάκλιση για τον DTW εμφανίζονται με διακεκομμένη μαύρη γραμμή. Η θέση του μέγιστου αρμονικού μέσου φαίνεται με μαύρο αστέρι. Εμείς επικυρώσαμε την έξοδο της συνάρτησης κόστους για διαφορετικές τιμές των κατωφλίων.

Μια σημαντική διαφορά μεταξύ των δύο μεθόδων COSFIRE και DTW είναι ο χρόνος που χρειάζεται για να εκτελέσει ένα συγκεκριμένο τραγούδι. Έχει αποδειχθεί ότι ο χρόνος της επεξεργασίας του DTW είναι σημαντικά υψηλότερος από τη μέθοδο COSFIRE.

Χρόνος Επεξεργασίας (s)		
Τραγούδι	COSFIRE	DTW
Gia tin kerinia	50	641
O pramateftis	67	770
Afkoritissa	12	133
Nanourisma 1	38	532
Nanourisma 2	20	196
Nanourisma 3	18	202
Nanourisma 4	55	897
Tis sousas 1	5	561
Tis sousas 2	9	192

Πίνακας 6.2. Χρόνος επεξεργασίας σε δευτερόλεπτα για τις δύο μεθόδους για εννέα τραγούδια.

Κεφάλαιο 7

Συμπεράσματα και Μελλοντική Εργασία

7.1 Συμπεράσματα	34
7.2 Μελλοντική Εργασία	35

7.1 Συμπεράσματα

Μετά την ανάλυση των αποτελεσμάτων που λήφθηκαν από την όλη διαδικασία που ακολουθήθηκε, καταλήξαμε σε ορισμένα συμπεράσματα. Πετύχαμε αρκετά καλά αποτελέσματα θέτοντας έτσι βάσεις για τυχόν μελλοντικές εργασίες.

Για αυτή την Ατομική Διπλωματική Εργασία, έγινε μια ανασκόπηση της βιβλιογραφίας, πιο συγκεκριμένα ανάλυση της μουσικής με υπολογιστικές μεθόδους όπως η τεχνικές μηχανικής μάθησης, εξαγωγή χαρακτηριστικών κτλ. Στη συνέχεια έγινε υλοποίηση διάφορων τεχνικών με την πλατφόρμα MATLAB, καθώς και μια προεπεξεργασία των δεδομένων με σκοπό να κατανοηθεί καλύτερα η βάση δεδομένων και να γίνει πιο αποδοτική για το σύστημα που θα υλοποιήσουμε.

Ταυτοχρόνως με τα πιο πάνω, δημιουργήθηκε μια πλατφόρμα στο διαδίκτυο όπου συμπεριλαμβάνονται όλα τα σχετικά της Ατομικής Διπλωματικής μου Εργασίας. Καθ' όλη τη διάρκεια αυτού του έτους σπουδών μου, έγιναν συναντήσεις με τον καθηγητή μου αλλά και τον βοηθό μου, Ανδρέα Νεοκλέους σε τακτικά χρονικά διαστήματα με σκοπό την ενημέρωση της δουλειάς μου και την καθοδήγηση από τους επιβλέποντες μου για την ολοκλήρωση της Ατομικής Διπλωματικής μου Εργασίας. Ακόμη ένας από τους στόχους μου ήταν να τελειώσω στην ώρα μου αυτή την εργασία σύμφωνα με το χρονοδιάγραμμα.

Τώρα όσον αφορά τη δουλειά που έγινε, στο κεφάλαιο 6 παρουσιάσαμε τα αποτελέσματα για τον COSFIRE και τον DTW λαμβάνοντας υπόψη την ακρίβεια (Precision - P) και την ανάκληση (Recall - R). Οι διαφορετικές τιμές της P και της R για τον COSFIRE λήφθηκαν με την αλλαγή της παραμέτρου «άλφα». Χρησιμοποιήσαμε 17 διαφορετικές τιμές της παραμέτρου «άλφα» στο εύρος [0,1 – 0,9]. Η P μειώνεται και η R αυξάνεται με την αύξηση των τιμών του «άλφα». Από τις τιμές αυτές καταλήγουμε στο συμπέρασμα ότι η μέθοδος COSFIRE αποδίδει καλύτερα από τη μέθοδο DTW.

Ένα άλλο σημαντικό αποτέλεσμα που λάβαμε μεταξύ των δύο μεθόδων COSFIRE και DTW είναι ο χρόνος που χρειάζεται για να εκτελέσει ένα συγκεκριμένο τραγούδι. Έχει αποδειχθεί ότι ο χρόνος της επεξεργασίας του DTW είναι σημαντικά υψηλότερος από τη μέθοδο COSFIRE, άρα συμπεραίνουμε ότι εκτός από την απόδοση στις πιο πάνω τιμές, ο COSFIRE είναι και πιο ευέλικτος από τον αλγόριθμο DTW.

Τέλος, κλείνοντας με τα συμπεράσματα, παρατηρούμε ότι η μέθοδος που περιγράψαμε στο κεφάλαιο 5.3.2. για την εύρεση των παύσεων μεταξύ των φωνητικών φράσεων, δεν είναι πολύ αποτελεσματική όταν χρησιμοποιείται σε διάφορα άλλα ηχητικά σήματα, αλλά μόνο για τις δικές μας περιπτώσεις που έχουμε αναφέρει. Ο λόγος είναι ότι το κατώφλι που χρησιμοποιούμε είναι σταθερό, ενώ πρέπει να είναι διαφορετικό για κάθε τραγούδι. Σημειωτέο ότι προσπάθειες για ανεύρεση μιας τέτοιας μεθόδου ξεκίνησαν και συνεχίζονται εδώ και χρόνια. Μια μέθοδος που θα μπορούσε να χρησιμοποιηθεί είναι με τη χρήση ιστογραμμάτων, όπου το κατώφλι θα είναι μεταβλητό για κάθε τραγούδι.

7.2 Μελλοντική Εργασία

Φθάνοντας εις πέρας αυτής της Ατομικής Διπλωματικής Εργασίας διαπιστώσαμε ότι η υλοποίηση τέτοιων συστημάτων μάθησης και η προσπάθεια δημιουργίας ενός αλγόριθμου που να αναγνωρίζει επαναλήψεις στη μελωδία ενός τραγουδιού, βοηθά σε διάφορους τομείς, όπως της Μουσικής, της Επιστήμης των Υπολογιστών καθώς επίσης και άλλων. Για τον λόγο αυτό πιστεύω ότι καλό θα ήταν να μην σταματήσει εδώ αυτή η μελέτη που έγινε, αλλά αντιθέτως να επεκταθεί μελλοντικά. Έχοντας ως υπόβαθρο την δουλειά που προσκόμισα σε αυτή τη Ατομική Διπλωματική Εργασία, θα αναφέρω κάποιες προτάσεις βασισμένες σ' αυτή, που

πιστεύω ότι αξίζει να μελετηθούν και να υλοποιηθούν έτσι ώστε να υποβοηθηθούν ακόμη περισσότεροι τομείς στον κλάδο που ονομάζεται Επιστήμη της Πληροφορικής. Συνοπτικά αυτές οι προτάσεις φαίνονται και πιο κάτω:

- Εισαγωγή χαρακτηριστικών για γρηγορότερη εκτέλεση του αλγορίθμου
- Δημιουργία γραφικής διεπαφής χρήστη
- Εισαγωγή και άλλων τραγουδιών εκτός από τα Κυπριακά λαϊκά τραγούδια
- Ανάπτυξη ή αντικατάσταση μεθόδου κατάτμησης σήματος σε μοτίβα
- Βελτίωση αλγορίθμου COSFIRE

Πιο λεπτομερώς:

Μια πρώτη ιδέα για μια πιθανή μελλοντική εργασία μπορεί να αφορά την εισαγωγή και άλλων χαρακτηριστικών, όπως για παράδειγμα την κλίση της καμπύλης, την ενέργεια σήματος, την απόσταση διαδοχικών δειγμάτων κτλ. με σκοπό την γρηγορότερη εκτέλεση του αλγορίθμου. Έτσι με αυτόν τον τρόπο ίσως αυξηθούν και βελτιστοποιηθούν και τα αποτελέσματα. Ακόμη μια πιθανή επέκταση του συστήματος είναι η δημιουργία γραφικής διεπαφής χρήστη για ευκολότερη χρήση. Η εισαγωγή και άλλων τραγουδιών εκτός από τα Κυπριακά λαϊκά τραγούδια αποτελεί μια άλλη πιθανή προσέγγιση για μελλοντική μελέτη. Έπειτα, στο παρόν σύστημα μας απαιτείται κατάτμηση του σήματος σε μοτίβα. Ωστόσο, αυτή η μέθοδος δεν είναι και πολύ αξιόπιστη και χρειάζεται ανάπτυξη ή αντικατάσταση από μια άλλη πιο έμπιστη μέθοδο. Τέλος, βελτίωση χρειάζεται και ο αλγόριθμος COSFIRE, έτσι ώστε να μην χρειάζεται η κατάτμηση των μοτίβων που αναφέραμε.

Βιβλιογραφία

- [1] G. Tzanetakis, A. Kapur, W. A. Schloss, M. Wright “Computational Ethnomusicology, journal of interdisciplinary music studies”, 2007.
- [2] O. Lartillot, “MIRtoolbox” ,Finnish Centre of Excellence in Interdisciplinary Music Research Swiss Center for Affective Sciences, 2013.
- [3] G. Azzopardi and N. Petkov, Trainable COSFIRE filters for keypoint detection and pattern recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, in print, 2012.
- [4] Bohak C., Marolt M., “Finding repeating stanzas in folk songs”, 13th International society for music information retrieval conference(ISMIR), 2012.
- [5] P.Van Kranenburg and G.Tzanetakis. A computational approach to the modeling and employment of cognitive units of folk song melodies using audio recordings. In Proceedings of the 11th International Conference on Music Perception and Cognition,2010.
- [6] Dr. R. Schreiber, “Matlab” , Hewlett-Packard Labs, Scholarpedia, 2007.
- [7] I.H. Witten, E.Frank, M.A. Hall , “Data Mining” , The University Of Waikatom, 2011.
- [8] L.R Rabiner, R.W. Schafer, “Introduction to Digital Speech Processing”, Rutgers University and University of California, Santa Barbara, USA and Hewlett-Packard Laboratories, Palo Alto, CA, USA, 2007.

- [9] R.N. Bracewell, “The Fourier transform and its applications”, Stanford University, 1999.
- [10] P. Domingos “A Few Useful Things to Know about Machine Learning”, Department of Computer Science and Engineering University Of Washington, 2000.
- [11] K.A.D. Jong, W.M. Spears, D.F. Gordon , “Using Genetic Algorithms for Concept Learning”, Naval Research Laboratory, Washington, 1993.
- [12] J.R. Quinlan, “The ID3 Algorithm”, University Of Sydney, 1997.
- [13] K. Chai , H. L. Chieu “Bayesian Online Classifiers for Text Classification and Filtering”, Proceedings of the 25th annual international ACM SIGIR conference on Research and Development in Information Retrieval, 2002.
- [14] H.A Rowley, S. Baluja, T. Kanade, “Neural Network-Based Detection”, 1998.
- [15] R. Agrawal, R. Srikant “Fast algorithms for mining association rules in large databases”, Proceedings of the 20th International Conference on Very Large Data Bases, 1994.
- [16] L.P.. Kaelbling, M.L. Littman, A.W. Moore, “Reinforcement Learning: A Survey” , Computer Science Department, Brown University, 1996.
- [17] A.De Cheveigne and H.Kawahara. YIN, a fundamental frequency estimator for speech and music, The Journal of the Acoustical Society of America,2002.
- [18] M.Muller. Information Retrieval for Music and Motion. Springer Verlag, 200
- [19] M. A. Bartsch and G. H. Wakefield. Audio thumbnailing of popular music using chroma-based representations. IEEE Transactions on Multimedia, 7(1):96–104, Feb. 2005.