

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ



ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ
**«ΑΝΑΓΝΩΡΙΣΗ ΦΥΣΙΚΩΝ ΠΡΟΣΩΠΩΝ ΜΕ ΒΑΣΗ ΤΑ ΗΧΗΤΙΚΑ
ΣΗΜΑΤΑ ΟΜΙΛΙΑΣ ΚΑΙ ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ»**

Ανδρέας Πολίτης
ΑΜ: i1398
Επιβλέπων: Γιαννακέας Νικόλαος

Άρτα, Νοέμβριος 2020

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ



ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ:

**«ΑΝΑΓΝΩΡΙΣΗ ΦΥΣΙΚΩΝ ΠΡΟΣΩΠΩΝ ΜΕ ΒΑΣΗ ΤΑ ΗΧΗΤΙΚΑ
ΣΗΜΑΤΑ ΟΜΙΛΙΑΣ ΚΑΙ ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ»**

Ανδρέας Πολίτης

ΑΜ: i1398

Επιβλέπων Καθηγητής: Γιαννακέας Νικόλαος

- Άρτα 2020-

**« IDENTIFICATION OF NATURALS BASED ON
SPEECH SOUNDS AND USE OF MECHANICAL
LEARNING METHODS »**

Εγκρίθηκε από τριμελή εξεταστική επιτροπή

Άρτα, Απρίλιος 2020

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Επιβλέπων καθηγητής

Νικόλαος Γιαννακέας,

Επίκουρος Καθηγητής

2. Μέλος επιτροπής

Τζάλλας Αλέξανδρος,

Επίκουρος Καθηγητής

3. Μέλος επιτροπής

Δημήτριος Δημόπουλος,

Πανεπιστημιακός Υπότροφος

Ο Προϊστάμενος του Τμήματος

Ευριπίδης Γλαβάς,

Καθηγητής, Α' Βαθμίδας

Υπογραφή

Δήλωση μη λογοκλοπής

Δηλώνω υπεύθυνα και γνωρίζοντας τις κυρώσεις του Ν. 2121/1993 περί Πνευματικής Ιδιοκτησίας, ότι η παρούσα πτυχιακή εργασία είναι εξ ολοκλήρου αποτέλεσμα δικής μου ερευνητικής εργασίας, δεν αποτελεί προϊόν αντιγραφής ούτε προέρχεται από ανάθεση σε τρίτους. Όλες οι πηγές που χρησιμοποιήθηκαν (κάθε είδους, μορφής και προέλευσης) για τη συγγραφή της περιλαμβάνονται στη βιβλιογραφία.

Ανδρέας Πολίτης

Υπογραφή

ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστώ θερμά τον καθηγητή μου κ. Γιαννακέα Νικόλαο για την άψογη συνεργασία που είχαμε, τον χρόνο που μου διέθεσε και την σωστή καθοδήγηση σε όλη τη διάρκεια της παρούσας πτυχιακής εργασίας.

Περίληψη

Η εργασία αφορά στην αναγνώριση ενός φυσικού προσώπου από την ομιλία του, μεταξύ πολλών συμμετεχόντων σε διάλογο. Η αναγνώριση θα επιτυγχάνεται με χρήση εκπαιδευμένου συστήματος και αλγορίθμων μηχανικής μάθησης. Πιο συγκεκριμένα, θα γίνει καταγραφή ομιλίας διαφορετικών ατόμων, το ηχητικό σήμα των οποίων θα επεξεργασθεί και θα εξαχθούν χαρακτηριστικά του σήματος. Το σύνολο των χαρακτηριστικών θα εκπαιδεύσουν ευφυής αλγορίθμους ταξινόμησης με σκοπό κάθε να καταστεί δυνατός ο διαχωρισμός των προσώπων που ομιλούν συγκεκριμένες χρονικές στιγμές. Ένα σύστημα με τα χαρακτηριστικά που περιγράφουμε θα μπορούσε να εφαρμοστεί για παράδειγμα σε περιβάλλοντα υποβοηθούμενη διαβίωσης (assisted living) με σκοπό την αναγνώριση του ατόμου που ενδεχομένως χρειάζεται βοήθεια μόνο από την ομιλία του.

Λέξεις-κλειδιά: Ομιλία, Αλγόριθμος μηχανικής μάθησης, Ηχητικό σήμα

Abstract

The work concerns the identification of a natural person from his speech, among many participants in a dialogue. Recognition will be achieved using a trained system and machine learning algorithms. More specifically, the speech of different people will be recorded, the audio signal of which will be processed and the characteristics of the signal will be extracted. The set of features will train intelligent classification algorithms in order to make it possible to distinguish between persons speaking at specific times. A system with the characteristics we describe could be applied for example in assisted living environments in order to identify the person who may need help only from his speech.

Keywords: Talk, Machine learning algorithm, Sound signal

Περιεχόμενα

Δήλωση μη λογοκλοπής	7
ΕΥΧΑΡΙΣΤΙΕΣ	8
Περίληψη	9
Abstract.....	10
ΚΕΦΑΛΑΙΟ 1 ΑΝΑΛΥΣΗ ΗΧΗΤΙΚΩΝ ΣΗΜΑΤΩΝ	12
1.1 Ηχητικά Σήματα.....	12
1.2 Ψηφιακή Αναπαράσταση Ηχητικών Σημάτων.....	13
1.2.1 Συχνότητες Ηχητικών Σημάτων	15
1.2.2 Καταγραφές Ηχητικών Σημάτων (μικρόφωνα, αρχές λειτουργίας κτ'λ)	20
1.3 Ηχητικά Σήματα Ομιλίας	21
1.3.1. Φυσιολογία Παραγωγής Φωνής (φωνητικές χορδές κτλ).....	21
1.3.2 Χαρακτηριστικά Ηχητικών Σημάτων Ομιλίας (συχνότητες, db κτλ)	22
ΚΕΦΑΛΑΙΟ 2 ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ ΟΜΙΛΙΑΣ	24
2.1 Προεπεξεργασία Σήματος Ομιλίας (αναλογική – ψηφιακή μετατροπή).....	24
2.2 Εφαρμογές Επεξεργασίας Ομιλίας. (π.χ. αναγνώριση φυσικής γλώσσας).....	27
2.3 Έρευνες που έχουν Γίνει (5-6 papers ή αναγνώριση ατόμων από την ομιλία τους).....	29
ΚΕΦΑΛΑΙΟ 3 ΥΠΟΔΟΜΗ	32
3.1 Το Πρόγραμμα Ανοιχτού Κώδικα OpenVibe	32
3.1.1 Σύγκριση με Άλλες Πλατφόρμες BCI	33
3.1.2 Διαφορετικοί τύποι χρηστών	34
3.1.3 Σχεδιασμός BCI με το OpenViBE	35
3.2 Το Λογισμικό Weka για Εξόρυξη Δεδομένων.....	38
ΚΕΦΑΛΑΙΟ 4 ΜΕΘΟΔΟΛΟΓΙΑ	40
4.1 Μηχανική Μάθηση και data mining.....	40
4.2 Σύνολο Δεδομένων	44
4.4 Μέθοδοι Ταξινόμησης (Περιγραφή Αλγορίθμων)	44
4.4.1 Linear Discriminant analysis (LDA)	44
4.4.2 Νευρωνικά δίκτυα (Perceptron)	45
4.4.3 Μηχανές Διανυσμάτων υποστήριξης (Support Vector Machines - SVM).....	47
4.4.4 Δένδρα Απόφασης – (Περιγραφή Αλγορίθμου C4.5)	48

ΚΕΦΑΛΑΙΟ 1 ΑΝΑΛΥΣΗ ΗΧΗΤΙΚΩΝ ΣΗΜΑΤΩΝ

1.1 Ηχητικά Σήματα

Η επεξεργασία σήματος ήχου είναι ένα υποπεδίο επεξεργασίας σήματος που σχετίζεται με τον ηλεκτρονικό χειρισμό ηχητικών σημάτων. Τα ηχητικά σήματα είναι ηλεκτρονικές αναπαραστάσεις ηχητικών κυμάτων - διαμήκη κύματα που ταξιδεύουν μέσω του αέρα, αποτελούμενα από συμπίεση και σπάνιες. Η ενέργεια που περιέχεται στα ηχητικά σήματα μετράτε συνήθως σε ντεσιμπέλ. Καθώς τα ηχητικά σήματα μπορούν να αναπαρασταθούν σε ψηφιακή ή αναλογική μορφή, η επεξεργασία μπορεί να πραγματοποιηθεί σε οποιονδήποτε τομέα. Οι αναλογικοί επεξεργαστές λειτουργούν απευθείας στο ηλεκτρικό σήμα, ενώ οι ψηφιακοί επεξεργαστές λειτουργούν μαθηματικά στην ψηφιακή του αναπαράσταση (Hodgson, J, 2010).

Ένα ηχητικό σήμα είναι μια αναπαράσταση του ήχου, συνήθως χρησιμοποιώντας είτε ένα επίπεδο ηλεκτρικής τάσης για αναλογικά σήματα, είτε μια σειρά δυαδικών αριθμών για ψηφιακά σήματα. Τα ηχητικά σήματα έχουν συχνότητες στο εύρος συχνοτήτων ήχου περίπου 20 έως 20.000 Hz, το οποίο αντιστοιχεί στα χαμηλότερα και ανώτερα όρια της ανθρώπινης ακοής. Τα ηχητικά σήματα μπορούν να συντίθενται άμεσα ή μπορεί να προέρχονται από έναν μορφοτροπέα όπως ένα μικρόφωνο, ένα μουσικό όργανο, ένα φυσίγγιο φωνογράφου ή μια κεφαλή ταινίας. Τα ηχεία ή τα ακουστικά μετατρέπουν ένα ηλεκτρικό σήμα ήχου σε ήχο (Atti, Andreas Spanias, 2006).

Τα ψηφιακά συστήματα ήχου αντιπροσωπεύουν ηχητικά σήματα σε μια ποικιλία ψηφιακών μορφών (Atti, Andreas Spanias, 2006). Ένα κανάλι ήχου είναι ένα κανάλι επικοινωνίας σήματος ήχου σε μια συσκευή αποθήκευσης ή μια κονσόλα μίξης, που χρησιμοποιείται σε λειτουργίες όπως εγγραφή πολλαπλών κομματιών και ενίσχυση ήχου.

1.2 Ψηφιακή Αναπαράσταση Ηχητικών Σημάτων

Ο ψηφιακός ήχος είναι μια αναπαράσταση του ήχου που καταγράφεται ή μετατρέπεται σε ψηφιακή μορφή. Στον ψηφιακό ήχο, το ηχητικό κύμα του ηχητικού σήματος κωδικοποιείται ως αριθμητικά δείγματα σε συνεχή ακολουθία. Για παράδειγμα, σε CD ήχου, τα δείγματα λαμβάνονται 44100 φορές το δευτερόλεπτο το καθένα με βάθος δείγματος 16 bit. Ο ψηφιακός ήχος είναι επίσης το όνομα για ολόκληρη την τεχνολογία ηχογράφησης και αναπαραγωγής χρησιμοποιώντας ηχητικά σήματα που έχουν κωδικοποιηθεί σε ψηφιακή μορφή. Μετά από σημαντικές εξελίξεις στην τεχνολογία ψηφιακού ήχου κατά τη διάρκεια της δεκαετίας του 1970, αντικατέστησε σταδιακά την αναλογική τεχνολογία ήχου σε πολλούς τομείς της μηχανικής ήχου και των τηλεπικοινωνιών τη δεκαετία του 1990 και του 2000 (Janssens, Jelle 2009).

Σε ένα ψηφιακό σύστημα ήχου, ένα αναλογικό ηλεκτρικό σήμα που αντιπροσωπεύει τον ήχο μετατρέπεται με έναν αναλογικό σε ψηφιακό μετατροπέα (ADC) σε ψηφιακό σήμα, συνήθως χρησιμοποιώντας διαμόρφωση παλμικού κώδικα. Αυτό το ψηφιακό σήμα μπορεί στη συνέχεια να εγγραφεί, να επεξεργαστεί, να τροποποιηθεί και να αντιγραφεί χρησιμοποιώντας υπολογιστές, μηχανήματα αναπαραγωγής ήχου και άλλα ψηφιακά εργαλεία. Όταν ο μηχανικός ήχου επιθυμεί να ακούσει την εγγραφή στα ακουστικά ή τα μεγάφωνα (ή όταν ένας καταναλωτής επιθυμεί να ακούσει ένα ψηφιακό αρχείο ήχου), ένας μετατροπέας ψηφιακού σε αναλογικό (DAC) εκτελεί την αντίστροφη διαδικασία, μετατρέποντας ξανά ένα ψηφιακό σήμα σε ένα αναλογικό σήμα, το οποίο στη συνέχεια αποστέλλεται μέσω ενός ενισχυτή ισχύος ήχου και τελικά σε ένα μεγάφωνο (Schroeder, Manfred R 2014).

Τα ψηφιακά συστήματα ήχου μπορεί να περιλαμβάνουν στοιχεία συμπίεσης, αποθήκευσης, επεξεργασίας και μετάδοσης. Η μετατροπή σε ψηφιακή μορφή επιτρέπει τον κατάλληλο χειρισμό, αποθήκευση, μετάδοση και ανάκτηση ενός ηχητικού σήματος. Σε αντίθεση με τον αναλογικό ήχο, στον οποίο η παραγωγή αντιγράφων μιας εγγραφής οδηγεί σε απώλεια παραγωγής και υποβάθμιση της ποιότητας του σήματος,

ο ψηφιακός ήχος επιτρέπει την άπειρο αριθμό αντιγράφων για τη δημιουργία χωρίς υποβάθμιση της ποιότητας του σήματος (Gray, Robert M 2010).

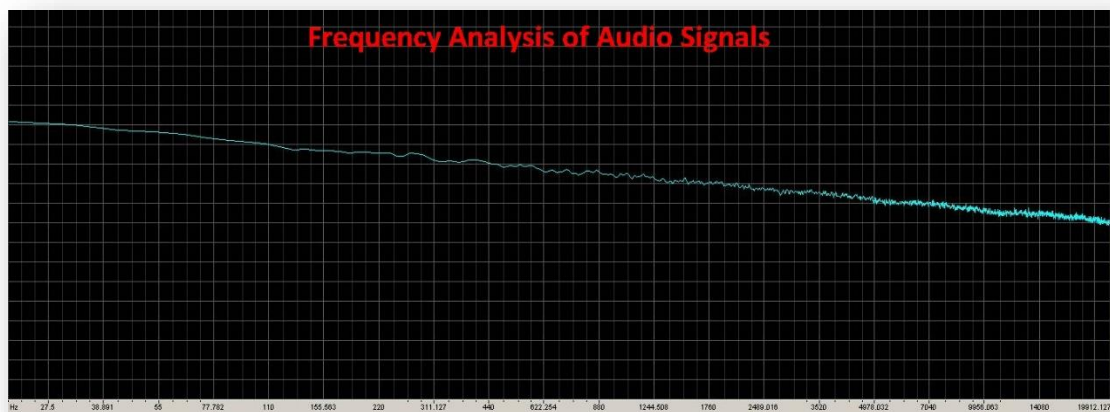
Οι ψηφιακές τεχνολογίες ήχου χρησιμοποιούνται για την εγγραφή, τον χειρισμό, τη μαζική παραγωγή και τη διανομή του ήχου, συμπεριλαμβανομένων ηχογραφήσεων τραγουδιών, οργάνων, podcast, ηχητικών εφέ και άλλων ήχων. Η σύγχρονη διαδικτυακή διανομή μουσικής εξαρτάται από την ψηφιακή εγγραφή και τη συμπίεση δεδομένων. Η διαθεσιμότητα της μουσικής ως αρχείων δεδομένων και όχι ως φυσικών αντικειμένων έχει μειώσει σημαντικά το κόστος διανομής (Janssens, Jelle, 2009). Πριν από τον ψηφιακό ήχο, η μουσική βιομηχανία διανέμει και πούλησε μουσική με την πώληση φυσικών αντιγράφων με τη μορφή δίσκων και κασετών. Με συστήματα ψηφιακού ήχου και διανομής όπως το iTunes, οι εταιρείες πωλούν ψηφιακά αρχεία ήχου στους καταναλωτές, τα οποία λαμβάνει ο καταναλωτής μέσω του Διαδικτύου.

Ένα αναλογικό σύστημα ήχου μετατρέπει τις φυσικές κυματομορφές του ήχου σε ηλεκτρικές αναπαραστάσεις αυτών των κυματομορφών χρησιμοποιώντας έναν μορφοτροπέα, όπως ένα μικρόφωνο. Οι ήχοι στη συνέχεια αποθηκεύονται σε αναλογικό μέσο όπως μαγνητική ταινία, ή μεταδίδονται μέσω αναλογικού μέσου όπως τηλεφωνικής γραμμής ή ραδιοφώνου. Η διαδικασία αντιστρέφεται για αναπαραγωγή: το ηλεκτρικό σήμα ήχου ενισχύεται και στη συνέχεια μετατρέπεται σε φυσικές κυματομορφές μέσω ενός ηχείου. Ο αναλογικός ήχος διατηρεί τα θεμελιώδη χαρακτηριστικά που μοιάζουν με κύματα σε όλη την αποθήκευση, τον μετασχηματισμό, την αναπαραγωγή και την ενίσχυση (Gray, Robert M 2010).

Τα αναλογικά ηχητικά σήματα είναι επιρρεπή σε θόρυβο και παραμόρφωση, λόγω των εγγενών χαρακτηριστικών των ηλεκτρονικών κυκλωμάτων και των σχετικών συσκευών. Οι διαταραχές σε ένα ψηφιακό σύστημα δεν οδηγούν σε σφάλμα εκτός εάν η διαταραχή είναι τόσο μεγάλη ώστε να έχει ως αποτέλεσμα να παρερμηνεύεται ένα σύμβολο ως άλλο σύμβολο ή να διαταράσσεται η ακολουθία των συμβόλων. Είναι επομένως γενικά δυνατό να υπάρχει ένα εντελώς ψηφιακό σύστημα ήχου χωρίς σφάλματα στο οποίο δεν εισάγεται θόρυβος ή παραμόρφωση μεταξύ μετατροπής σε ψηφιακή μορφή και μετατροπής σε αναλογική (Allstot, David J. 2016).

Ένα ψηφιακό σήμα ήχου μπορεί προαιρετικά να κωδικοποιηθεί για διόρθωση τυχόν σφαλμάτων που ενδέχεται να προκύψουν κατά την αποθήκευση ή μετάδοση του σήματος. Αυτή η τεχνική, γνωστή ως κωδικοποίηση καναλιού, είναι απαραίτητη για τη μετάδοση ή την εγγραφή ψηφιακών συστημάτων για τη διατήρηση της ακρίβειας bit. Η διαμόρφωση οκτώ έως δεκατέσσερα είναι ένας κωδικός καναλιού που χρησιμοποιείται στο CD ήχου (CD) (Floyd, Michael D. 2018).

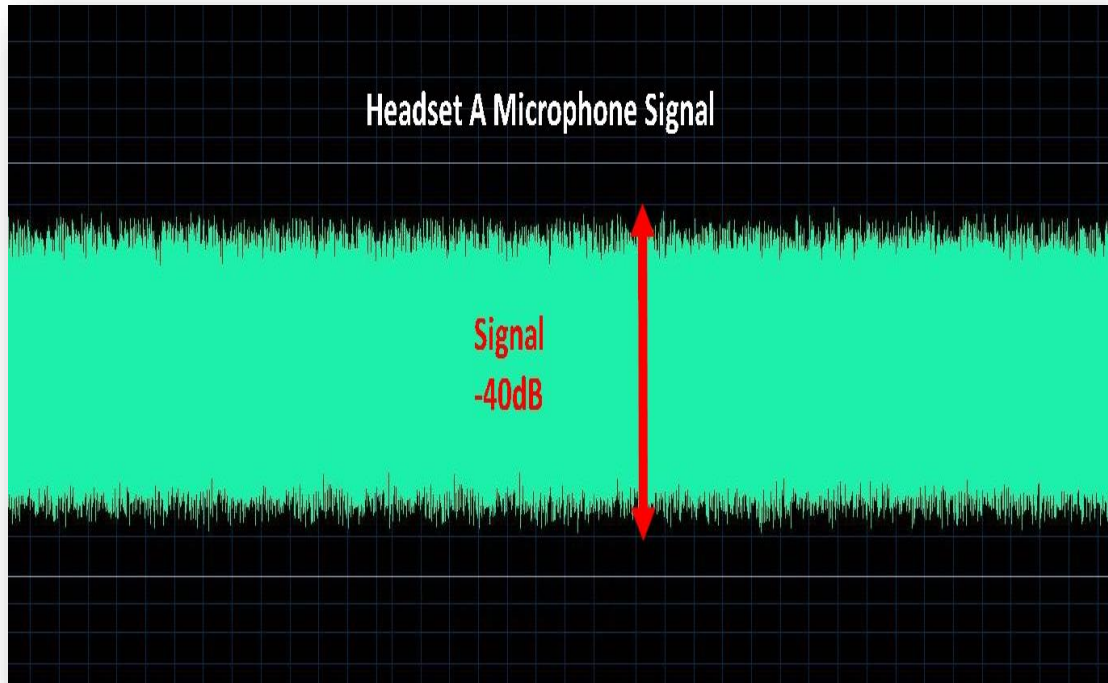
1.2.1 Συχνότητες Ηχητικών Σημάτων



Η Ανάλυση Συχνότητας των ηχητικών σημάτων λαμβάνει τα ηχητικά σήματα που είναι επιθυμητά να εξεταστούν και τα τοποθετεί στον τομέα συχνοτήτων για να παρατηρήσει πώς τα σήματα ανταποκρίνονται σε όλες τις διάφορες συχνότητες ενός δεδομένου εύρους ζώνης.

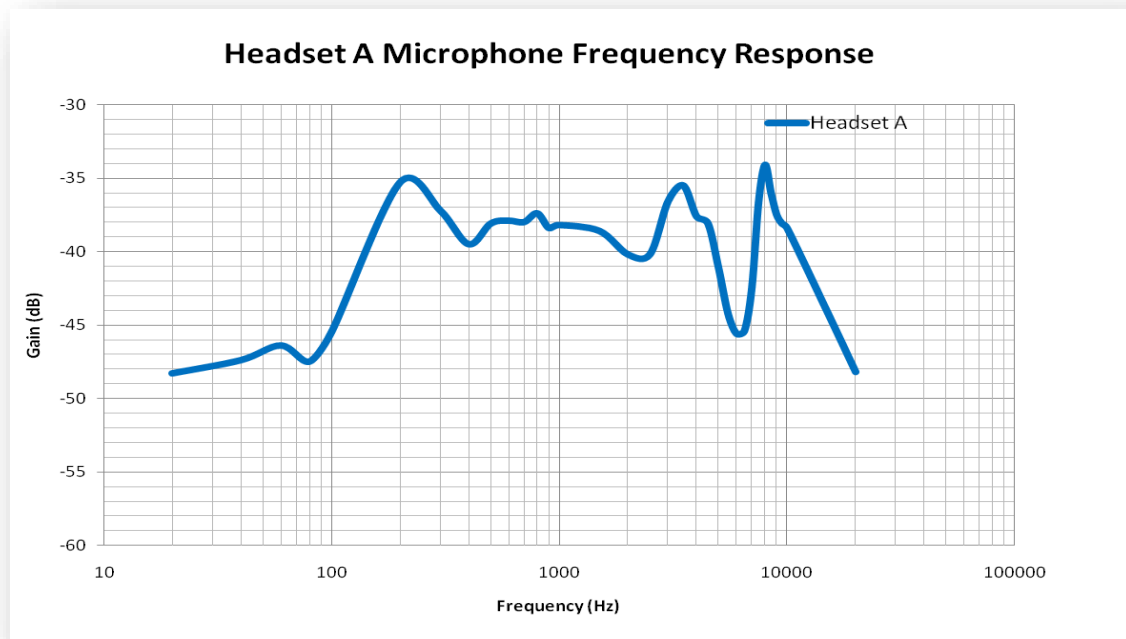
Συνήθως όταν ένα σήμα καταγράφεται τακτικά, είναι στον τομέα χρόνου. Για να το μετατρέψετε σε τομέα συχνότητας, πρέπει να χρησιμοποιηθεί ένα ειδικό λογισμικό ήχου που μπορεί να πραγματοποιήσει ανάλυση συχνότητας. Ο χρήστης μπορεί στη συνέχεια να καθορίσει σε ποια ζώνη ή εύρος συχνοτήτων θέλει να εξετάσει το ηχητικό σήμα (Wu J, 2012).

Σχήμα 1.2.1.1 Ακολουθεί ένα ηχητικό σήμα που καταγράφεται στον τομέα χρόνου
Πηγή: (Wu J, 2012).



Εδώ, όταν ένα ηχητικό σήμα βρίσκεται στον τομέα χρόνου, μπορείτε να δείτε μόνο το πλάτος του σήματος και τίποτα άλλο. Αυτό το πλάτος είναι το εύρος RMS, ή με πιο κατανοητούς όρους, το μέσο πλάτος του σήματος κατά μέσο όρο σε όλες τις διάφορες συχνότητες για το εύρος ζώνης που αναφέρεται.

Σχήμα 1.2.1.2 Τώρα παρακάτω είναι το ηχητικό σήμα στον τομέα συχνότητας Πηγή: (Wu J, 2012).



Εκτελώντας ανάλυση συχνότητας στο σήμα θορύβου που καταγράφει ένα μικρόφωνο είναι σημαντικό για να ανακαλύψετε ποιες συχνότητες ένα μικρόφωνο ακυρώνει τον θόρυβο πιο αποτελεσματικά και ποιες συχνότητες ακυρώνει τον θόρυβο πολύ λιγότερο.

Έτσι, με τη χρήση αυτής της μεθόδου, ένας μηχανικός μπορεί να ανακαλύψει ποιες συχνότητες ακουστικά ακυρώνει τον θόρυβο με τον πιο φτωχό και στη συνέχεια μπορεί να επανασχεδιάσει το ακουστικό έτσι ώστε να ακυρώνει καλύτερα τον θόρυβο σε αυτές τις συχνότητες. Έτσι, μπορεί να βελτιώσει τη συνολική ικανότητα ακύρωσης θορύβου του ακουστικού (Lippert M, 2007).

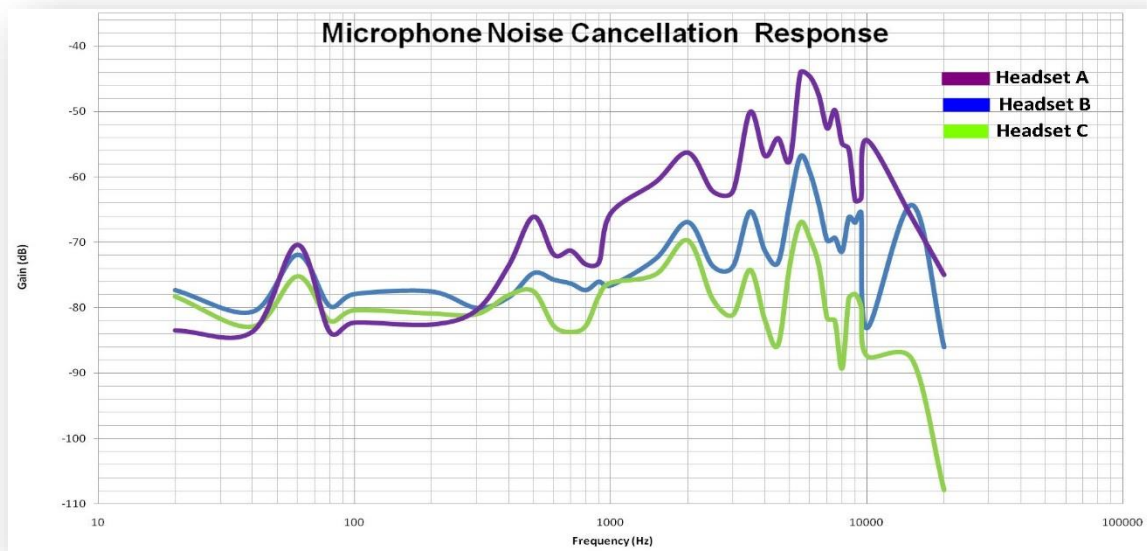
Σχήμα 1.2.1.3 Ακολουθεί ένα γράφημα που δείχνει τα σήματα θορύβου που καταγράφηκαν από τρία ακουστικά Πηγή: (Wu J, 2012).



Όλα τα ακουστικά εκτέθηκαν στο ίδιο ακριβώς επίπεδο θορύβου ροζ. Ο ροζ θόρυβος προσομοιώνει τον θόρυβο στο παρασκήνιο που τα ακουστικά πρέπει να ακυρώσουν. Έτσι, το ακουστικό που καταγράφει το ελάχιστο ποσό θορύβου έχει την καλύτερη ακύρωση. Μπορούμε να δούμε ότι το ακουστικό C έχει την ακύρωση θορύβου και το ακουστικό A έχει τη χειρότερη ακύρωση με το ακουστικό B στο μεταξύ.

Τώρα παρακάτω είναι το γράφημα της απόκρισης συχνότητας των ακουστικών. Το παραπάνω γράφημα είναι το σήμα θορύβου που εμφανίζεται σε απόκριση του χρόνου, αλλά τώρα το παρακάτω γράφημα είναι τα σήματα θορύβου ως απόκριση στη συχνότητα. Έτσι, μπορούμε να δούμε πώς τα μικρόφωνα στα ακουστικά ανταποκρίνονται συγκεκριμένα σε κάθε συχνότητα εντός του δοκιμασμένου εύρους ζώνης μας, που είναι 20Hz έως 20KHz (.Yang W, 2013).

Σχήμα 1.2.1.4 καθαρότητα ακουστικού : (Wu J, 2012).



Με βάση αυτό το γράφημα, μπορούμε να δούμε καθαρά ποιο ακουστικό έχει την καλύτερη ακύρωση θορύβου. Το ακουστικό C καταγράφει το λιγότερο σήμα, οπότε είναι πάνω ακυρώνει τον μεγαλύτερο θόρυβο ενώ το ακουστικό A λαμβάνει το μεγαλύτερο σήμα, οπότε έχει τη χειρότερη ακύρωση. Ωστόσο, αυτός δεν είναι ο λόγος για τον οποίο εξετάζουμε το γράφημα ανάλυσης συχνότητας των ακουστικών, γιατί θα μπορούσαμε να πούμε ποιο ακουστικό έχει την καλύτερη ακύρωση θορύβου κοιτάζοντας απλώς το γράφημα στον τομέα χρόνου (Yang W, 2013).

Ο λόγος για τον οποίο εξετάζουμε συγκεκριμένα ένα γράφημα ανάλυσης συχνότητας είναι έτσι ώστε να μπορούμε να δούμε ακριβώς ποιες συχνότητες ακουστικά ακυρώνουν τον καλύτερο θόρυβο και ποιες συχνότητες ακουστικά ακυρώνουν τον θόρυβο το χειρότερο. Ως επί το πλείστον, όλα τα ακουστικά ακυρώνουν τις χαμηλές συχνότητες καλύτερα από τις υψηλές συχνότητες με πολύ περισσότερα. Ας ρίξουμε μια ματιά τώρα ειδικά στο μωβ ακουστικό, το Ακουστικό A. Μπορείτε να δείτε τη δραστική διαφορά στα επίπεδα στο πώς ακυρώνει τις χαμηλές συχνότητες έναντι των υψηλών συχνοτήτων. Για πολλές από τις χαμηλές συχνότητες από περίπου 20Hz-45Hz και 80Hz-300Hz, το Headset C έχει την καλύτερη ακύρωση από όλα τα ακουστικά. Ωστόσο, στα μεσαία και στις υψηλές συχνότητες, από 300Hz έως περίπου 15KHz, έχει τη χειρότερη ακύρωση, καθώς καταγράφει τον περισσότερο θόρυβο από όλα τα άλλα ακουστικά.

Έτσι, ένας μηχανικός θα ήξερε ότι θα έπρεπε να παράγει καλύτερη ακύρωση θορύβου στις συχνότητες από 300Hz έως 15KHz για να παράγει την ακύρωση θορύβου του ακουστικού C. Αυτός είναι ο τρόπος με τον οποίο η ανάλυση συχνότητας είναι σημαντική (Pérez-Bellido A, 2013).

Εδώ μπορείτε να δείτε τι πλάτος παράγει το σήμα ήχου σε κάθε συγκεκριμένη συχνότητα. Ενώ το εύρος περιοχής χρόνου σας δίνει μόνο το εύρος RMS κατά μέσο όρο σε όλες τις διάφορες συχνότητες, η εξέταση του ηχητικού σήματος στον τομέα συχνοτήτων σας επιτρέπει να βλέπετε κάθε μεμονωμένο πλάτος για κάθε συχνότητα.

Η ανάλυση συχνότητας είναι σημαντική γιατί είναι σημαντικό να γνωρίζουμε πώς οι συσκευές ήχου ενισχύουν τον ήχο σε συγκεκριμένες συχνότητες. Τα ηχεία και τα μικρόφωνα πρέπει να ενισχύουν ορισμένους ήχους ανάλογα με την εφαρμογή που χρησιμοποιείται. Για παράδειγμα, ένα ηχείο μπορεί να χρειαστεί να αναπαράγει ήχους υψηλής συχνότητας, οπότε πρέπει να είναι σε θέση να ενισχύσει τον ήχο σε υψηλές συχνότητες. Επομένως, είναι σημαντικό να εκτελέσετε ανάλυση συχνότητας στο ηχείο για να βεβαιωθείτε ότι συμβαίνει αυτό (Pérez-Bellido A, 2013).

1.2.2 Καταγραφής Ηχητικών Σημάτων (μικρόφωνα, αρχές λειτουργίας κτ'λ)

Η ηχογράφηση και η αναπαραγωγή είναι μια ηλεκτρική, μηχανική, ηλεκτρονική ή ψηφιακή επιγραφή και επαναδημιουργία ηχητικών κυμάτων, όπως ομιλούμενη φωνή, τραγούδι, οργανική μουσική ή ηχητικά εφέ. Οι δύο κύριες κατηγορίες τεχνολογίας ηχογράφησης είναι η αναλογική εγγραφή και η ψηφιακή εγγραφή (Koetsier, 2001).

Η ακουστική αναλογική εγγραφή επιτυγχάνεται με ένα διάφραγμα μικροφώνου που ανιχνεύει αλλαγές στην ατμοσφαιρική πίεση που προκαλούνται από ακουστικά ηχητικά κύματα και τα καταγράφει ως μηχανική αναπαράσταση των ηχητικών κυμάτων σε ένα μέσο όπως μια φωνογραφική εγγραφή (στην οποία μια γραφίδα κόβει αυλάκια σε μια εγγραφή) . Στην μαγνητική μαγνητοταινία, τα ηχητικά κύματα δονούν το διάφραγμα του μικροφώνου και μετατρέπονται σε ένα μεταβαλλόμενο ηλεκτρικό ρεύμα, το οποίο στη συνέχεια μετατρέπεται σε ένα διαφορετικό μαγνητικό πεδίο από έναν ηλεκτρομαγνήτη, το οποίο κάνει μια αναπαράσταση του ήχου ως μαγνητισμένες περιοχές σε μια πλαστική ταινία με μαγνητική επικάλυψη σε αυτό. Η αναλογική

αναπαραγωγή ήχου είναι η αντίστροφη διαδικασία, με μεγαλύτερο διάφραγμα ηχείων που προκαλεί αλλαγές στην ατμοσφαιρική πίεση για να σχηματίσουν ακουστικά ηχητικά κύματα (Mitchell, 2006).

Η ψηφιακή εγγραφή και αναπαραγωγή μετατρέπει το αναλογικό σήμα ήχου που λαμβάνεται από το μικρόφωνο σε ψηφιακή μορφή μέσω της διαδικασίας δειγματοληψίας. Αυτό επιτρέπει στα δεδομένα ήχου να αποθηκεύονται και να μεταδίδονται από μια ευρύτερη ποικιλία μέσων. Η ψηφιακή εγγραφή αποθηκεύει τον ήχο ως μια σειρά δυαδικών αριθμών (μηδενικά και αυτά) που αντιπροσωπεύουν δείγματα του πλάτους του ηχητικού σήματος σε ίσα χρονικά διαστήματα, με ρυθμό δειγμάτων αρκετά υψηλό για να μεταφέρει όλους τους ήχους που μπορούν να ακουστούν. Ένα ψηφιακό σήμα ήχου πρέπει να μετατραπεί σε αναλογική μορφή κατά την αναπαραγωγή πριν ενισχυθεί και συνδεθεί σε ένα μεγάφωνο για να παράγει ήχο.

Πριν από την ανάπτυξη της ηχογράφησης, υπήρχαν μηχανικά συστήματα, όπως κουτιά μουσικής εκκαθάρισης και, αργότερα, πιάνο αναπαραγωγής, για κωδικοποίηση και αναπαραγωγή οργανικής μουσικής (Copeland, 2008).

1.3 Ηχητικά Σήματα Ομιλίας

1.3.1. Φυσιολογία Παραγωγής Φωνής (φωνητικές χορδές κτλ)

Τα ανθρώπινα όργανα παραγωγής ομιλίας είναι πολλαπλών χρήσεων: δημιουργία λόγου, αναπνοή, φαγητό και αίσθηση οσμών. Έτσι, από την άποψη της επικοινωνίας, η παραγωγή ομιλίας δεν μπορεί να είναι η βέλτιστη πηγή πληροφοριών, καθώς το αντί είναι ένας δέκτης. Ορισμένες παραλληλισμοί μπορούν να γίνουν μεταξύ ηλεκτρονικής και ανθρώπινης ομιλίας. Οι άνθρωποι ελαχιστοποιούν την προσπάθεια, από την άποψη της ενέργειας και του χρόνου τους, ενώ μεγιστοποιούν την αντιληπτική αντίθεση στους ακροατές (Pitsikalis, V., 2006).

Φωνητική οδός

Οι πνεύμονες παρέχουν τη ροή του αέρα και την πηγή πίεσης για ομιλία, και τα φωνητικά κορδόνια συνήθως ρυθμίζουν τη ροή του αέρα για να δημιουργούν πολλές παραλλαγές ήχου. Ωστόσο, είναι η φωνητική οδός που είναι το πιο σημαντικό

συστατικό του συστήματος στην παραγωγή ανθρώπινης ομιλίας. Είναι μια δίοδος που μοιάζει με σωλήνα που αποτελείται από μυς και άλλους ιστούς και επιτρέπει την παραγωγή των διαφορετικών ήχων που αποτελούν την προφορική γλώσσα (Pitsikalis, V., 2006).

Για τους περισσότερους ήχους, η φωνητική οδός τροποποιεί τη χρονική και φασματική κατανομή ισχύος στα ηχητικά κύματα, τα οποία ξεκινούν από τη γλωττίδα. Επιπλέον, το φωνητικό κομμάτι παράγει μερικούς ήχους απευθείας. Είναι η πηγή για παρεμπόδιση όπως η στάση και οι τριβείς ήχοι. Διαφορετικά τηλέφωνα μπορούν κυρίως να διακριθούν από την περιοδικότητά τους - φωνητικά ή μη φωνητικά, φασματικό σχήμα - σχεδόν ποιες συχνότητες έχουν μεγάλη ισχύ και διάρκεια - τα μεγαλύτερα τηλέφωνα θεωρούνται ότι έχουν μεγαλύτερο άγχος. Η κατάσταση των φωνητικών πτυχών καθορίζει συνήθως την επιλογή λειτουργιών φωνής κάθε τηλεφώνου (Potamianos, G., 2004). Η πιο σημαντική πτυχή της παραγωγής λόγου είναι μακράν η προδιαγραφή διαφορετικών τηλεφώνων μέσω των ενεργειών φιλτραρίσματος της φωνητικής οδού. Επειδή η αντίληψη του λόγου κυριαρχείται από την παρουσία της ηχητικής ισχύος. Οι σχηματιστές συντομογραφούνται συχνά F1, επομένως F1 σημαίνει το σχηματιστή με τη χαμηλότερη συχνότητα. Στα φωνητικά τηλέφωνα, οι σχηματιστές μειώνουν συχνά την ισχύ ως συνάρτηση της συχνότητας λόγω της γενικής διέγερσης της γλωττίδας της φύσης lowpass, επομένως το F1 είναι συνήθως το ισχυρότερο φορμά. Η μετατόπιση των αρθρωτών αλλάζει το σχήμα του ακουστικού σωλήνα μέσω του οποίου περνά ο ήχος και μεταβάλλει την απόκριση συχνότητας. Μετά την έξοδο από τον λάρυγγα, ο αέρας από τους πνεύμονες περνά μέσα από τις φάρυγγες και τις στοματικές κοιλότητες και τα εξανθήματα στα χείλη. Για ρινικούς ήχους, ο αέρας αφήνεται να εισέλθει στη ρινική κοιλότητα κατεβάζοντας το βαλμό, στο όριο μεταξύ των φάρυγγων και των στοματικών κοιλοτήτων (Potamianos, G., 2004).

1.3.2 Χαρακτηριστικά Ηχητικών Σημάτων Ομιλίας (συχνότητες, db κτλ)

Η ομιλία είναι μια κυματομορφή πίεσης που ταξιδεύει από έναν ομιλητή σε έναν ή περισσότερους ακροατές. Αυτό το σήμα μετράται συνήθως ακριβώς μπροστά από το στόμα του ηχείου, το οποίο είναι η κύρια θέση εξόδου για την ομιλία. Δεδομένου ότι η ατμόσφαιρα περιβάλλοντος στην οποία κάποιος μιλά επιβάλλει μια βασική πίεση, στην πραγματικότητα είναι η διακύμανση της πίεσης που προκαλείται από το ηχείο που

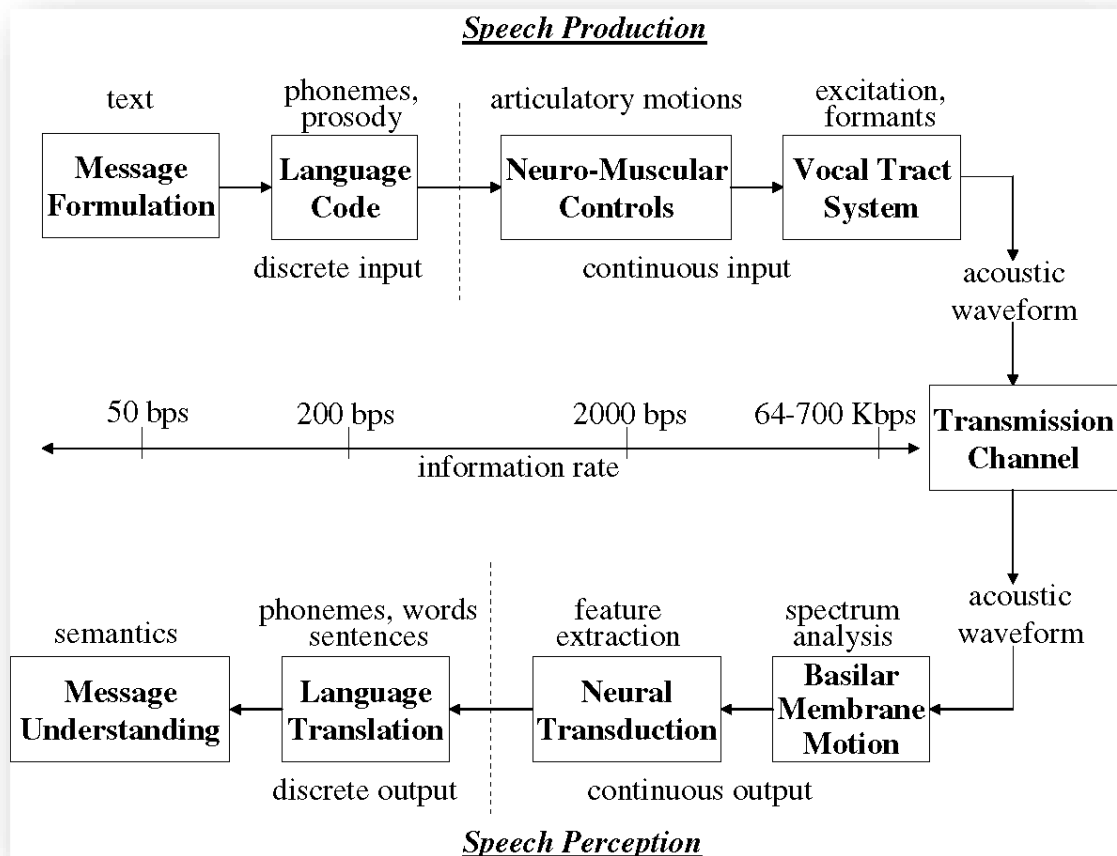
αποτελεί το σήμα ομιλίας. Το σήμα είναι συνεχές στη φύση και είναι πολύ δυναμικό σε χρόνο και πλάτος, που αντιστοιχεί στη συνεχώς μεταβαλλόμενη κατάσταση του φωνητικού συστήματος και των φωνητικών κορδονιών (Anderson, T.W 2003). Η ομιλία μπορεί να χαρακτηριστεί ως διακριτή ακολουθία τμημάτων ήχου που ονομάζονται τηλέφωνα, καθένα από τα οποία έχει ορισμένες ακουστικές και αρθρωτικές ιδιότητες κατά τη σύντομη χρονική περίοδο. Κάθε φωνή επιβάλλει ορισμένους περιορισμούς στις θέσεις αυτών των αρθρωτών ή οργάνων του φωνητικού συστήματος: φωνητικές πτυχές ή φωνητικά κορδόνια, γλώσσα, χείλη, δόντια, βαλμό και γνάθο. Οι ήχοι ομιλίας εμπίπτουν σε δύο ευρείες τάξεις: (α) φωνήεντα που είναι υπεύθυνα να επιτρέπουν απεριόριστη ροή αέρα σε ολόκληρη τη φωνητική οδό και (β) σύμφωνα με τα οποία ελέγχουν τη ροή σε κάποιο σημείο και έχουν πιο αδύναμα για τα φωνήεντα (Anderson, T.W 2003).

ΚΕΦΑΛΑΙΟ 2 ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ ΟΜΙΛΙΑΣ

2.1 Προεπεξεργασία Σήματος Ομιλίας (αναλογική – ψηφιακή μετατροπή)

Το σχήμα 2.1.1 δείχνει την πλήρη διαδικασία παραγωγής και αντίληψης ομιλίας από τη διατύπωση ενός μηνύματος στον εγκέφαλο ενός ομιλητή, προς τη δημιουργία του σήματος ομιλίας και, τέλος, στην κατανόηση του μηνύματος από έναν ακροατή. Στην κλασική εισαγωγή τους στην επιστήμη της ομιλίας, οι Denes και Pinson αναφέρθηκαν κατάλληλα σε αυτήν τη διαδικασία ως αλυσίδα λόγου (Denes 1993). Η διαδικασία ξεκινά επάνω αριστερά καθώς ένα μήνυμα αντιπροσωπεύεται κάπως στον εγκέφαλο του ομιλητή. Οι πληροφορίες μηνυμάτων μπορούν να θεωρηθούν ότι έχουν έναν αριθμό διαφορετικών αναπαραστάσεων κατά τη διάρκεια της διαδικασίας παραγωγής ομιλίας.

Για παράδειγμα, το μήνυμα θα μπορούσε να αναπαρασταθεί αρχικά ως αγγλικό κείμενο. Για να μιλήσει στο μήνυμα, ο ομιλητής μετατρέπει έμμεσα το κείμενο σε συμβολική αναπαράσταση της ακολουθίας ήχων που αντιστοιχεί στην εκφωνούμενη έκδοση του κειμένου. Αυτό το βήμα, που ονομάζεται γεννήτρια κωδικών γλώσσας στο Σχήμα 2.1.1, μετατρέπει τα σύμβολα κειμένου σε φωνητικά σύμβολα (μαζί με πληροφορίες στρες και διάρκειας) που περιγράφουν τους βασικούς ήχους μιας εκφωνούμενης έκδοσης του μηνύματος και τον τρόπο (δηλαδή, την ταχύτητα και την έμφαση) που προορίζονται να παράγονται οι ήχοι.



Σχήμα 2.1.1 Η αλυσίδα Ομιλίας.

Για παράδειγμα, τα τμήματα της κυματομορφής του Σχήματος 2.1.1 επισημαίνονται με φωνητικά σύμβολα χρησιμοποιώντας έναν κωδικό φιλικό προς το πληκτρολόγιο υπολογιστή που ονομάζεται ARPAbet.1 Έτσι, το κείμενο θα πρέπει να κυνηγούμε αντιπροσωπεύεται φωνητικά (σε σύμβολα ARPAbet) ως [SH UH Δ - Δ IY - CH EY S]. Το τρίτο βήμα στη διαδικασία παραγωγής ομιλίας είναι η μετατροπή σε νευρομυϊκούς ελέγχους, δηλαδή, το σύνολο των σημάτων ελέγχου που κατευθύνουν το νευρο-μυϊκό σύστημα για να μετακινήσουν τους ρυθμιστές ομιλίας, δηλαδή τη γλώσσα, τα χείλη, τα δόντια, και το σαγόι, με τρόπο που να συμβαδίζει με τους ήχους του επιθυμητού προφορικού μηνύματος και με τον επιθυμητό βαθμό έμφασης. Το τελικό αποτέλεσμα του βήματος των νευρομυϊκών μαρτύρων είναι ένα σύνολο αρθρικών κινήσεων (συνεχής έλεγχος) που αναγκάζουν τους αρθρωτές της φωνητικής οδού να κινηθούν με καθορισμένο τρόπο προκειμένου να δημιουργήσουν τους επιθυμητούς ήχους.

Τέλος, το τελευταίο βήμα στη διαδικασία Παραγωγής Ομιλίας είναι το φωνητικό σύστημα συστήματος που δημιουργεί φυσικά τις απαραίτητες πηγές ήχου και τα κατάλληλα σχήματα φωνητικής οδού με την πάροδο του χρόνου, έτσι ώστε να δημιουργείται μια ακουστική κυματομορφή, όπως αυτή που φαίνεται στο σχήμα 2.1.1, που κωδικοποιεί τις πληροφορίες στο επιθυμητό μήνυμα στο σήμα ομιλίας.

Για τον προσδιορισμό του ρυθμού ροής πληροφοριών κατά την παραγωγή ομιλίας, υποθέστε ότι υπάρχουν περίπου 32 σύμβολα (γράμματα) στη γλώσσα (στα Αγγλικά υπάρχουν 26 γράμματα, αλλά αν συμπεριλαμβάνουμε απλή στίξη πλησιάζουμε $32 = 25$ σύμβολα). Επιπλέον, το ποσοστό μιλώντας για τους περισσότερους ανθρώπους είναι περίπου 10 σύμβολα ανά δευτερόλεπτο (κάπως στην υψηλή πλευρά, αλλά εξακολουθεί να είναι αποδεκτή για μια πρόβλεψη κατά προσέγγιση ρυθμού πληροφοριών). Ως εκ τούτου, υποθέτοντας ανεξάρτητα γράμματα ως απλή προσέγγιση, εκτιμούμε το βασικό ρυθμό πληροφοριών του μηνύματος κειμένου ως περίπου 50 bps (5 bit ανά σύμβολο φορές 10 σύμβολα ανά δευτερόλεπτο). Στο δεύτερο στάδιο της διαδικασίας, όπου η αναπαράσταση κειμένου μετατρέπεται σε δείκτες φωνημάτων (π.χ., βήμα και άγχος), ο ρυθμός πληροφοριών εκτιμάται ότι αυξάνεται κατά ένα συντελεστή 4 έως περίπου 200 bps. Για παράδειγμα, το σύνολο φωνητικών συμβόλων ARBAbet που χρησιμοποιείται για την επισήμανση των ήχων ομιλίας στο Σχήμα 2.2.1 περιέχει περίπου $64 = 26$ σύμβολα ή περίπου 6 bit / φωνήμα (πάλι μια πρόχειρη προσέγγιση που προϋποθέτει ανεξαρτησία των φωνημάτων). Στο Σχήμα 2.1.1, υπάρχουν 8 φωνήματα σε περίπου 600 ms. Αυτό οδηγεί σε εκτίμηση $8 \times 6 / 0,6 = 80$ bps. Πρόσθετες πληροφορίες που απαιτούνται για να περιγράψουν προσωδιακά χαρακτηριστικά του σήματος (π.χ. διάρκεια, βήμα, ένταση) θα μπορούσαν εύκολα να προσθέσουν 100 bps στο συνολικό ρυθμό πληροφοριών για ένα μήνυμα που κωδικοποιείται ως σήμα ομιλίας.

Οι αναπαραστάσεις πληροφοριών για τα δύο πρώτα στάδια της ομιλίας η αλυσίδα είναι διακριτή, ώστε να μπορούμε να εκτιμήσουμε εύκολα το ποσοστό των πληροφοριών, που ρέει με μερικές απλές υποθέσεις. Για το επόμενο στάδιο της ομιλίας το μέρος παραγωγής της αλυσίδας ομιλίας, η αναπαράσταση γίνεται συνεχής (με τη μορφή σημάτων ελέγχου για αρθρωτική κίνηση). Εάν μπορούσαν να μετρηθούν, θα μπορούσαμε να εκτιμήσουμε το φασματικό εύρος ζώνης αυτών των σημάτων ελέγχου και να δείξουμε και να ποσοτικοποιήσουμε κατάλληλα αυτά τα σήματα για να λάβουμε ισοδύναμα ψηφιακά σήματα για τα οποία μπορεί να εκτιμηθεί ο ρυθμός δεδομένων.

Οι αρθρωτές κινούνται σχετικά αργά σε σύγκριση με το χρόνο παραλλαγής της προκύπτουσας ακουστικής κυματομορφής. Οι Εκτιμήσεις εύρους ζώνης και η απαιτούμενη ακρίβεια υποδηλώνει ότι ο συνολικός ρυθμός δεδομένων του δείγματος τα αρθρωτικά σήματα ελέγχου είναι περίπου 2000 bps (Flanagan 1972). Έτσι, το πρωτότυπο μήνυμα κειμένου αντιπροσωπεύεται από ένα σύνολο συνεχώς μεταβαλλόμενων σημάτων, του οποίου η ψηφιακή αναπαράσταση απαιτεί πολύ υψηλότερο ρυθμό δεδομένων από το ρυθμό πληροφοριών που εκτιμήσαμε για τη μετάδοση του μηνύματος ως σήμα ομιλίας. Τέλος, όπως θα δούμε αργότερα, ο ρυθμός δεδομένων της ψηφιοποιημένης κυματομορφής ομιλίας στο τέλος του τμήματος παραγωγής ομιλίας της αλυσίδας ομιλίας μπορεί να είναι οπουδήποτε από 64.000 σε περισσότερα από 700.000 bps. Φτάνουμε σε αυτούς τους αριθμούς εξετάζοντας το ρυθμό δειγματοληψίας και τον ποσοτικό προσδιορισμό που απαιτούνται για να αντιπροσωπεύσουμε το σήμα ομιλίας με την επιθυμητή αντιληπτή πιστότητα. Για παράδειγμα, η ποιότητα τηλεφώνου απαιτεί τη διατήρηση ενός εύρους ζώνης 0–4 kHz, υπονοώντας ένα ρυθμό δειγματοληψίας 8000 δειγμάτων / δευτ. Κάθε δείγμα μπορεί να κβαντιστεί με 8 bit σε μια κλίμακα καταγραφής, με αποτέλεσμα ένα ρυθμό bit 64.000 bps. Αυτή η αναπαράσταση είναι εξαιρετικά κατανοητή (δηλαδή, οι άνθρωποι μπορούν εύκολα να εξαγάγουν το μήνυμα από αυτό), αλλά στους περισσότερους ακροατές, θα ακούγεται διαφορετικό από το αρχικό σήμα ομιλίας που εκφώνησε ο ομιλητής. Από την άλλη πλευρά, η κυματομορφή ομιλίας μπορεί να αναπαρασταθεί με ποιότητα CD χρησιμοποιώντας ρυθμό δειγματοληψίας 44.100 δειγμάτων / δευτερόλεπτα με δείγματα 16 bit ή ρυθμό δεδομένων 705.600 bps. Σε αυτήν την περίπτωση, το αναπαραγόμενο ακουστικό σήμα θα είναι σχεδόν μη διακριτό από το αρχικό σήμα ομιλίας.

2.2 Εφαρμογές Επεξεργασίας Ομιλίας. (π.χ. αναγνώριση φυσικής γλώσσας)

Η αναγνώριση ομιλίας είναι μια περίπλοκη διαδικασία που αποτελείται από διαδοχικούς μετασχηματισμούς του σήματος ομιλίας που προηγούνται της εξαγωγής των πληροφοριακών χαρακτηριστικών του και διαφόρων επιπέδων αναγνώρισης που αποκωδικοποιούν τα φωνήματα, τις λέξεις, τις προτάσεις και, τελικά, την αίσθηση της πρότασης. Στο χαμηλότερο, φωνητικό επίπεδο αναγνώρισης, πραγματοποιείται η μετάβαση από το συνεχές κύμα ομιλίας στην ακολουθία διακριτών στοιχείων,

φωνημάτων. Ο αριθμός των φωνητικών διαφορών σε διαφορετικές γλώσσες και συνήθως κυμαίνεται σε αρκετές δεκάδες. Η κωδικοποίηση του σήματος ομιλίας σε μια ακολουθία φωνημάτων συνοδεύεται από ισχυρή συμπίεση πληροφοριών και, ως αποτέλεσμα, σημαντική μείωση στην περιγραφή του σήματος. Οι λέξεις, οι προτάσεις και, φυσικά, η αίσθηση της πρότασης αναγνωρίζονται βάσει της εξαγόμενης αλυσίδας φωνημάτων. Η διαδικασία αναγνώρισης φωνημάτων είναι απαραίτητο συστατικό όλων των υπάρχοντων συστημάτων σήματος ομιλίας που βασίζονται σε μεγάλα λεξικά που αποτελούνται από αρκετές χιλιάδες έως εκατοντάδες χιλιάδων λέξεων. Ωστόσο, η ποιότητα της αναγνώρισης φωνητικών παραμένει χαμηλή και το ποσοστό της σωστής αναγνώρισής τους δεν υπερβαίνει το 70 {80% σε ολόκληρο το αλφάβητο των φωνημάτων. Τα σφάλματα της κωδικοποίησης μέσω φωνητικού κώδικα διορθώνονται σε υψηλότερα επίπεδα αναγνώρισης χρησιμοποιώντας τις πληροφορίες που περιέχονται στο λεξικό του συστήματος αναγνώρισης και τα γλωσσικά μοντέλα. Ως αποτέλεσμα, κάποιος παίρνει την επιθυμητή ακρίβεια της αναγνώρισης λέξεων και προτάσεων εντός αρκετών τοις εκατό. Για παράδειγμα, εάν η πιθανότητα αναγνώρισης φωνημάτων είναι 0,8, η πιθανότητα αναγνώρισης μιας λέξης ve-phonem που λείπει στο λεξικό είναι περίπου 0,33. Η συγκριτικά χαμηλή ποιότητα της τηλεφωνικής κωδικοποίησης οφείλεται σε έναν αριθμό παραγόντων όπως η αξιοσημείωτη μεταβλητότητα στην προφορά του ηχείου, οι διαφορές μεταξύ των ηχείων, η ευαισθησία στον θόρυβο και οι παραμορφώσεις συχνότητας που εισήγαγε η ακουστική του δωματίου και οι αποκρίσεις συχνότητας του μικροφώνου και του καναλιού επικοινωνίας. Αυτοί οι παράγοντες έχουν ως αποτέλεσμα σημαντικές διακυμάνσεις στο φάσμα του σήματος ομιλίας που εμποδίζει τη σύγκριση του με τα φασματικά φωνήματα αναφοράς και προκαλεί τα σφάλματα αναγνώρισης. Η ανθρώπινη αντίληψη της ομιλίας, ωστόσο, είναι ασθενώς ευαίσθητη σε αυτούς τους παράγοντες και υπερβαίνει ουσιαστικά τα υπάρχοντα συστήματα αυτόματης αναγνώρισης ομιλίας σε ό, τι αφορά την ακρίβεια της ανάλυσης φωνημάτων. Αυτό φαίνεται να οφείλεται στο γεγονός ότι το ανθρώπινο ακουστικό σύστημα διαθέτει έναν πιο τέλειο μηχανισμό αναγνώρισης ομιλίας ικανό της εξαγωγής των κατάλληλων πληροφοριακών χαρακτηριστικών των φωνημάτων.

2.3 Έρευνες που έχουν Γίνει (5-6 papers ή αναγνώριση ατόμων από την ομιλίας τους

Η επεξεργασία σήματος ομιλίας είναι μια ενεργός ερευνητική περιοχή στον τομέα της επεξεργασίας ψηφιακού σήματος. Η επεξεργασία σήματος ομιλίας ήρθε στην εικόνα στη δεκαετία του '70. Μερικές από τις σημαντικές πτυχές της επεξεργασίας ψηφιακής ομιλίας είναι κωδικοποίηση υψηλής ποιότητας (αντιληπτική κωδικοποίηση) ομιλίας και ήχου, αναγνώριση ομιλίας, βελτίωση και τροποποίηση ομιλίας και ήχου, σύνθεση κειμένου σε ομιλία, σύνθεση ομιλίας σε κείμενο κ.λπ. σε αυτό το άρθρο εξετάζουμε τα τελευταία τέσσερα χρόνια στον τομέα της επεξεργασίας σήματος ομιλίας. Έχουμε εξετάσει τις τέσσερις εργασίες επεξεργασίας σήματος ομιλίας και όλες αυτές τις δημοσιεύσεις που δημοσιεύθηκαν από το 2017 έως το 2018. Τρεις δημοσιεύσεις δημοσιεύθηκαν το 2018 και μία δημοσιεύθηκε το 2017. Αυτές οι δημοσιεύσεις δημοσιεύονται σε περιοδικά και πρακτικά συνεδρίων του IEEE. Έτσι, η πλήρης διαδικασία αναθεώρησης βασίζεται στην πρόσφατη έρευνα σχετικά με την επεξεργασία σήματος ομιλίας από τα τελευταία τρία χρόνια.

II. ΑΝΑΣΚΟΠΗΣΗ ΧΑΡΤΩΝ

Ο Michał Raczynski et-al, 2018, περιγράφει έναν αλγόριθμο για την αναγνώριση του ήχου από προηγούμενες δημιουργημένες λέξεις. Χρησιμοποιούν το MATLAB για την εκτέλεση αυτής της εργασίας. Πρώτα περνούν το σήμα μέσω του φίλτρου. Καταγράφουν το σήμα με κάρτα μουσικής matlab για 1 δευτερόλεπτο. στη συνέχεια μετατρέψτε το αναλογικό σήμα σε ψηφιακό σήμα με συχνότητα δειγματοληψίας 8 khz. Το ληφθέν σήμα προσαρμόζεται σε σταθερό μήκος πλαισίου ίσο με 6000 δείγματα, χρησιμοποιώντας μέθοδο παρόμοια. Χρησιμοποιούν την τεχνική PLA για να αναγνωρίσουν τον κόσμο.

Ο Hariz Zakka Muhammad et al, το 2018 σχεδιάστηκε ως μοντέλο αναγνώρισης ομιλίας όπου η ομιλία ερμηνεύονταν σε ινδονησιακή γλώσσα. Χρησιμοποιούν τη μέθοδο εξαγωγής χαρακτηριστικών Mel Frequency Cepstral Coefficients (MFCC) και για το σκοπό αυτό χρησιμοποιήθηκαν το "Markov Model" Για το σκοπό αυτό χρησιμοποιούν την ακόλουθη διαδικασία διαγράμματος μπλοκ. Η πλήρης διαδικασία

χωρίστηκε σε δύο μέρη, η πρώτη εκπαίδευση και η δεύτερη είναι η διαδικασία δοκιμών.

Εφαρμόζουν τον αλγόριθμο ομαδοποίησης k-means στα επιλεγμένα δεδομένα. Είναι ένας αλγόριθμος χωρίς επίβλεψη. Η διαδικασία αναγνώρισης γίνεται με την εύρεση της τιμής πιθανότητας όλων των μοντέλων λ. Η διαδικασία MFCC έχει τα εξής βήματα:

- ❖ Προ-έμφαση
- ❖ Διαμόρφωση
- ❖ Παράθυρο
- ❖ Fast Fourier Transform (FFT)
- ❖ Φίλτρο τράπεζες
- ❖ Διακριτός μετασχηματισμός συνημίτων (DCT)

Τέλος, εφαρμόζουν το κρυφό μοντέλο markov για βελτιστοποίηση. Βρήκαν τη μέση ακρίβεια 70 τοις εκατό.

Prashant Upadhyaya et al, Μετρήστε την απόδοση του εργαλείου αναγνώρισης ομιλίας kaldi χρησιμοποιώντας την Τεχνική προσαρμογής ηχείων (SAT). Χρησιμοποίησαν την τεχνική μηχανικής εκμάθησης για τον σκοπό ανίχνευσης. Το χαρακτηριστικό ήταν ακουστική ισορροπία σύμφωνα με τα σύμβολα MFCC 1000 και PLP. Ινδουιστική πρόταση από το σώμα της AMUAV. Τα μοντέλα ακουστικής εκπαίδευσης χρησιμοποιούν το κρυφό μοντέλο Markov και το μοντέλο Gaussian. (NHM-GMM) και περιορισμένο βάρος μετατροπής State Transducer (WFST) βελτιώθηκε κατά 6,93%. Το μοντέλο αναγνώρισης ομιλίας στο Kaladi βασίζεται στην τελική κατάσταση ο μετατροπέας με την κύρια βιβλιοθήκη έχει γραφτεί σε υποστήριξη γλώσσας C ++ Ο ευέλικτος κώδικας είναι εύκολα κατανοητός δραστήρια. Το Kaladi Apache v2.0 υποστηρίζεται από άδεια, οι οποίες είναι κατάλληλες μεγάλες επιχειρήσεις κοινότητας χρηστών. Επιτυγχάνουν τη βελτίωση της αναγνώρισης κατά 7%.

Ο Gabor Kiss et al, 2017, σκέφτεται μια διαγλωσσική εξέταση που δίνεται σε σχέση με τις προσδοκίες προσδοκίας της αθλιότητας στη βάση της διαδικασίας του λόγου

Η εξέτασή μας πραγματοποιήθηκε σε 3 ευρωπαϊκές διαλέκτους: Γερμανικά, Ουγγρικά και Ιταλικά. Αυτές οι ακουστικές επιλογές ανατέθηκαν, ως φορέας πληροφοριών του

δείκτη, που σχετίζονται με τη σοβαρότητα της θλίψης με συγκρίσιμη ανεξάρτητη διάλεκτο. Πολλές μονοφωνικές και διαγλωσσικές δοκιμές κατευθύνθηκαν. Η τεχνική είναι ακόμη κατάλληλη για την πρόβλεψη της σοβαρότητας της απογοήτευσης μέσα στην παρουσία μιας διάλεκτου που δεν χρησιμοποιείται καθόλου μέσω της εκπαίδευσης του μοντέλου. Οι δοκιμές καταδεικνύουν ξεκάθαρα ότι η αναγνώριση της τριγλωσσικής θλίψης επιτυγχάνεται συχνά και θα πρέπει να είναι δυνατό να αναπτυχθεί μια προγραμματισμένη συμπτωματική συσκευή για την αθλιότητα της εργασίας των αναλυτών ή για την κατανόηση της προβολής, με έναν εξαιρετικά διαγλωσσικό τρόπο.

Η επεξεργασία σήματος ομιλίας είναι μια ενεργός ερευνητική περιοχή στον τομέα της επεξεργασίας ψηφιακού σήματος. Η επεξεργασία σήματος ομιλίας ήρθε στην εικόνα στη δεκαετία του '70. Εδώ σε αυτό το άρθρο εξετάζουμε την εργασία των τελευταίων τεσσάρων ετών στον τομέα της επεξεργασίας σήματος ομιλίας. Έχουμε εξετάσει τις τέσσερις εργασίες επεξεργασίας σήματος ομιλίας και όλες αυτές τις δημοσιεύσεις που δημοσιεύθηκαν από το 2017 έως το 2018. Τρεις δημοσιεύσεις δημοσιεύθηκαν το 2018 και μία εργασία δημοσιεύθηκε το 2017. Διαπιστώσαμε ότι και οι τέσσερις ερευνητές έχουν χρησιμοποιήσει τις τεχνικές μηχανικής εκμάθησης για τη βελτίωση ομιλίας και αναγνώριση γλώσσας. Έτσι μπορούμε να πούμε ότι η αναγνώριση εικόνας είναι ένα σημαντικό εργαλείο στον τομέα της ομιλίας και της επεξεργασίας κειμένου.

ΚΕΦΑΛΑΙΟ 3 ΥΠΟΔΟΜΗ

3.1 Το Πρόγραμμα Ανοιχτού Κώδικα OpenVibe Λειτουργίες OpenViBE

Το OpenViBE είναι μια δωρεάν πλατφόρμα λογισμικού ανοιχτού κώδικα για το σχεδιασμό, τη δοκιμή και τη χρήση των διεπαφών εγκεφάλου-υπολογιστή. Η πλατφόρμα αποτελείται από ένα σύνολο λειτουργικών μονάδων που μπορούν να ενσωματωθούν εύκολα και αποτελεσματικά στο σχεδιασμό του BCI τόσο για πραγματικές όσο και για εφαρμογές VR. Τα βασικά χαρακτηριστικά της πλατφόρμας είναι:

Αρθρωτότητα και επαναχρησιμοποίηση. Η πλατφόρμα μας είναι ένα σύνολο από ενότητες λογισμικού που έχουν καταργηθεί για την απόκτηση, την προ-επεξεργασία, την επεξεργασία και την οπτικοποίηση εγκεφαλικών δεδομένων, καθώς και για την αλληλεπίδραση με οθόνες εικονικής πραγματικότητας. Το OpenViBE είναι ένα λογισμικό γενικού σκοπού που σημαίνει ότι οι χρήστες είναι σε θέση να προσθέσουν εύκολα νέες ενότητες λογισμικού για να καλύψουν τις ανάγκες τους. Αυτό διασφαλίζεται χάρη στο concept box, ένα στοιχειώδες συστατικό που είναι υπεύθυνο για ένα κλάσμα ολόκληρου του αγωγού επεξεργασίας, το οποίο επιτρέπει την ανάπτυξη επαναχρησιμοποιήσιμων εξαρτημάτων, μειώνει το χρόνο ανάπτυξης και βοηθά στην γρήγορη επέκταση των λειτουργιών.

Διαφορετικοί τύποι χρηστών. Το OpenViBE έχει σχεδιαστεί για διαφορετικούς τύπους χρηστών:

Προγραμματιστές εικονικής πραγματικότητας, κλινικοί γιατροί, ερευνητές BCI κ.λπ. αντιμετωπίζονται οι διάφορες ανάγκες τους και προτείνονται διαφορετικά εργαλεία για καθένα από αυτά, ανάλογα με τις δεξιότητες προγραμματισμού και τις γνώσεις τους στις εγκεφαλικές διαδικασίες.

Φορητότητα. Η πλατφόρμα λειτουργεί ανεξάρτητα από τους διαφορετικούς στόχους λογισμικού και τις συσκευές υλικού. Περιλαμβάνει ένα αφηρημένο επίπεδο αναπαράστασης που επιτρέπει την εκτέλεση με διάφορες μηχανές απόκτησης, όπως EEG ή MEG. Μπορεί να εκτελεστεί σε λειτουργικά συστήματα Windows και Linux

και περιλαμβάνει επίσης διαφορετικές τεχνικές οπτικοποίησης δεδομένων. Τέλος, βασίζεται σε δωρεάν και φορητά λογισμικά (π.χ. GTK + 1, IT ++ 2, GSL3, VRPN4, GCC5).

Σύνδεση με VR. Το λογισμικό μας μπορεί να ενσωματωθεί σε εφαρμογές VR προηγμένης τεχνολογίας. Το OpenViBE λειτουργεί ως εξωτερικό περιφερειακό για κάθε είδους πραγματικό και εικονικό περιβάλλον. Εκμεταλλεύεται επίσης τις οθόνες VR χάρη στην ελαφριά αφαίρεση μιας βιβλιοθήκης διαχείρισης σκηνογραφίας, επιτρέποντας την απεικόνιση της εγκεφαλικής δραστηριότητας με ευανάγνωστο τρόπο ή την παροχή κινήτρων σε περιβάλλοντα εκπαίδευσης (π.χ. για νευροαπόδοση).

3.1.1 Σύγκριση με Άλλες Πλατφόρμες BCI

Σε σύγκριση με άλλα λογισμικά BCI, η πλατφόρμα OpenViBE εμφανίζεται ως εξαιρετικά αρθρωτή. Αντιμετωπίζει τις ανάγκες διαφορετικών τύπων χρηστών (θα πρέπει να είναι προγραμματιστές ή μη προγραμματιστές) και προτείνει μια φιλική προς το χρήστη γλώσσα γραφικών η οποία επιτρέπει στους μη προγραμματιστές να σχεδιάζουν ένα BCI χωρίς να γράφουν μία μόνο γραμμή κώδικα. Αντιθέτως, όλες οι άλλες πλατφόρμες BCI απαιτούν κάποιο βαθμό δεξιοτήτων προγραμματισμού για να σχεδιάσουν ένα νέο BCI σε πραγματικό χρόνο από το μηδέν. Επιπλέον, η αρθρωτότητά τους είναι πιο χονδροειδής (εκτός από το BioSig), περιορίζοντας έτσι το φάσμα των πιθανών σχεδίων.

Το OpenViBE είναι επίσης φορητό, ανεξάρτητο από το υλικό ή το λογισμικό και βασίζεται εξ ολοκλήρου σε δωρεάν λογισμικά ανοιχτού κώδικα. Συγκριτικά, μεταξύ άλλων πλατφορμών BCI σε πραγματικό χρόνο, μόνο το BioSig είναι πλήρως ανοιχτού κώδικα, αλλά το πακέτο rtsBCI που απαιτείται για το διαδικτυακό και πραγματικό χρόνο BCI απαιτεί Matlab / Simulink το οποίο είναι ένα μη-ελεύθερο και ιδιόκτητο λογισμικό.

Το OpenViBE προτείνει τη δημιουργία διαδικτυακών σεναρίων αυτόματα από την ανάλυση file. Τέλος, σε αντίθεση με άλλες πλατφόρμες, το OpenViBE είναι κατάλληλο για εφαρμογές VR καθώς παρέχει αρκετά ενσωματωμένα εργαλεία για το σχεδιασμό καινοτόμων οθονών VR και ανατροφοδότησης καθώς και για την εκτέλεση 3D απεικόνιση της δραστηριότητας του εγκεφάλου σε πραγματικό χρόνο. Επιπλέον, το

OpenViBE μπορεί επίσης να χρησιμοποιηθεί ως συσκευή για οποιαδήποτε εφαρμογή VR.

3.1.2 Διαφορετικοί τύποι χρηστών

Το OpenViBE έχει σχεδιαστεί για τέσσερις τύπους χρηστών. Από την πρώτη πλευρά, ο προγραμματιστής και ο προγραμματιστής εφαρμογών είναι και οι δύο προγραμματιστές, από την άλλη πλευρά ο συγγραφέας και ο χειριστής δεν χρειάζονται δεξιότητες προγραμματισμού.

Ο προγραμματιστής (προγραμματιστής) έχει τη δυνατότητα να προσθέσει νέες λειτουργίες και να δοκιμάσει τα δικά του κομμάτια λογισμικού στο OpenViBE. Για το σκοπό αυτό, το OpenViBE παραδίδεται με ένα πλήρες κιτ ανάπτυξης λογισμικού (SDK). Αυτό το SDK παρέχει πρόσβαση σε λειτουργίες σε διαφορετικά επίπεδα ανάλογα με την εργασία που πρέπει να πραγματοποιήσετε. Υπάρχουν δύο κύριες κατηγορίες προγραμματιστών. Πρώτον, οι προγραμματιστές του πυρήνα που βελτιώνουν και τροποποιούν τις λειτουργίες του πυρήνα. Δεύτερον, προγραμματιστές προσθηκών που δημιουργούν νέες πρόσθετες λειτουργικές μονάδες. Ο προγραμματιστής εφαρμογών (προγραμματιστής) χρησιμοποιεί το SDK για τη δημιουργία αυτόνομων εφαρμογών, χρησιμοποιώντας το OpenViBE ως βιβλιοθήκη. Τέτοιες εφαρμογές κυμαίνονται από νέα εργαλεία όπως το πρόγραμμα επεξεργασίας οπτικών σεναρίων που περιγράφεται στην ενότητα 6, έως τις εξωτερικές εφαρμογές VR με τις οποίες μπορεί να αλληλοεπιδράσει ο χρήστης BCI.

Ο συγγραφέας (μη προγραμματιστής) χρησιμοποιεί το πρόγραμμα επεξεργασίας οπτικών σεναρίων για να τακτοποιήσει τα υπάρχοντα πλαίσια για να σχηματίσουν ένα σενάριο. Διαμορφώνει αυτά τα κουτιά και το σενάριο για να δημιουργήσει ένα πλήρες, έτοιμο προς χρήση σύστημα BCI. Ο συγγραφέας γνωρίζει τα εσωτερικά της πλατφόρμας μας καθώς και τα συστήματα BCI και είναι εξοικειωμένος με τη βασική επεξεργασία σήματος. Γνωρίζει επίσης το παράδειγμα αλληλεπίδρασης που πρέπει να χρησιμοποιήσει.

Ωστόσο, δεν χρειάζεται ισχυρές δεξιότητες προγραμματισμού υπολογιστών, επειδή χρησιμοποιεί ειδικά εργαλεία για την εκτέλεση των καθηκόντων του (βλ. Ενότητα 6).

Ο χειριστής (μη προγραμματιστής) γενικά θα ήταν κλινικός ή επαγγελματίας (δεν είναι ειδικός υπολογιστών ούτε εμπειρογνώμονας OpenViBE). Είναι υπεύθυνος για τη χρήση και τη λειτουργία των προκατασκευασμένων σεναρίων του συγγραφέα. Τότε

τρέχει απλώς το σενάριο. Γνωρίζει πώς πρέπει και μπορεί να λειτουργεί το σύστημα BCI και παρακολουθεί την εκτέλεση του συστήματος BCI χάρη σε ειδικά εξαρτήματα οπτικοποίησης.

Έχει κατανόηση των νευροφυσιολογικών σημάτων και μπορεί να βοηθήσει τον χρήστη BCI να βελτιώσει τον έλεγχο του στο σύστημα BCI.

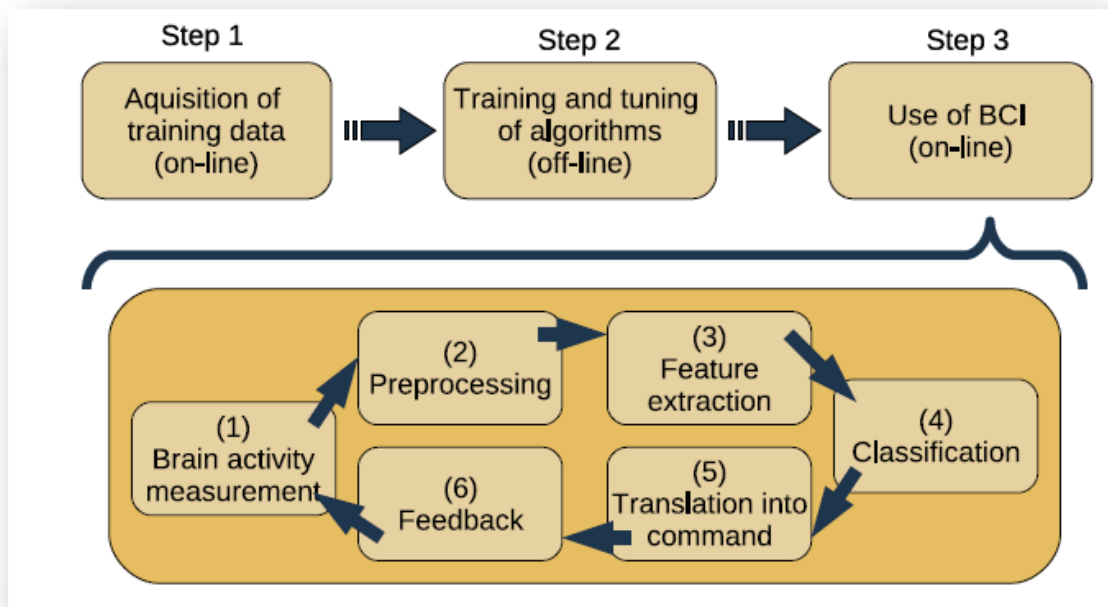
Τέλος, ένας άλλος ρόλος πρέπει να εξεταστεί: ο χρήστης BCI. Ο χρήστης BCI φορά γενικά το υλικό απόκτησης δραστηριότητας εγκεφάλου (π.χ. ένα καπάκι EEG) και αλληλοεπιδρά με μια εφαρμογή μέσω της ψυχικής του δραστηριότητας. Η εφαρμογή θα μπορούσε να είναι, για παράδειγμα, ένα πρόγραμμα εκπαίδευσης νευροανάδρασης, ένα βιντεοπαιχνίδι στην εικονική πραγματικότητα, μια απομακρυσμένη λειτουργία σε επαυξημένη πραγματικότητα κ.λπ. Ενώ δεν χρησιμοποιεί άμεσα την πλατφόρμα OpenViBE, εκμεταλλεύεται έμμεσα τις δυνατότητές της.

3.1.3 Σχεδιασμός BCI με το OpenViBE

Ο σχεδιασμός και η λειτουργία ενός διαδικτυακού BCI με το λογισμικό μας ακολουθεί έναν μάλλον καθολικό τρόπο για να το πράξει (Wolpaw et al., 2002). Απαιτούνται τρία ξεχωριστά βήματα (βλ. Εικόνα 3.1.3). Στο πρώτο βήμα, ένα σύνολο δεδομένων εκπαίδευσης πρέπει να καταγράφεται για ένα συγκεκριμένο θέμα, ενώ εκτελεί συγκεκριμένες ψυχικές εργασίες. Το δεύτερο βήμα συνίσταται σε μια ολοκληρωμένη ανάλυση αυτών των εγγραφών με στόχο την εύρεση των βέλτιστων παραμέτρων βαθμονόμησης

(π.χ. βέλτιστες δυνατότητες, σχετικά κανάλια κ.λπ.) για αυτό το θέμα. Το τελευταίο βήμα συνίσταται στη χρήση του BCI online σε μια διαδικασία κλειστού βρόχου. Προαιρετικά, οι επαναλήψεις μπορούν να γίνουν στην απόκτηση δεδομένων και στην εξειδικευμένη εκπαίδευση προκειμένου να προσδιοριστούν οι παράμετροι.

Ο διαδικτυακός βρόχος (τρίτο βήμα) είναι κοινός σε οποιοδήποτε BCI και αποτελείται από έξι φάσεις: μετρήσεις δραστηριότητας εγκεφάλου, προεπεξεργασία, εξαγωγή χαρακτηριστικών, ταξινόμηση, μετάφραση σε εντολή και ανατροφοδότηση (βλ. Σχήμα 3.1.3).



Εικόνα 3.1.3 Σχεδιάζοντας ένα BCI με OpenViBE

Μετρήσεις δραστηριότητας εγκεφάλου: Αυτό το βήμα συνίσταται στη μέτρηση της εγκεφαλικής δραστηριότητας του χρήστη BCI. Μέχρι σήμερα, περίπου μισή ντουζίνα διαφορετικά είδη εγκεφαλικών σημάτων έχουν αναγνωριστεί ως κατάλληλα για ένα BCI, δηλ., εύκολα παρατηρήσιμο και ελεγχόμενο (Wolpaw et al., 2002). Η μέτρηση της εγκεφαλικής δραστηριότητας για ένα σύστημα BCI πραγματοποιείται κυρίως χρησιμοποιώντας ηλεκτροεγκεφαλογραφία (EEG) καθώς είναι μια οικονομικά αποδοτική και μη επεμβατική μέθοδος που παρέχει υψηλή χρονική ανάλυση (Wolpaw et al., 2002). Το λογισμικό μας υποστηρίζει ήδη διάφορες συσκευές απόκτησης EEG, αλλά επίσης υποστηρίζει μηχανή μαγνητοεγκεφαλογραφίας (MEG).

Προ επεξεργασία: Το βήμα προ επεξεργασίας στοχεύει στην αποποίηση των ληφθέντων σημάτων και / ή στην ενίσχυση ενός συγκεκριμένου εγκεφαλικού σήματος (Bashashati, Fatourehchi, Ward, & Birch, 2007). Για παράδειγμα, το λογισμικό μας προτείνει διαφορετικά είδη αλγορίθμων προ επεξεργασίας όπως χρονικά φίλτρα και χωρικά φίλτρα (ανεξάρτητο στοιχείο, ανάλυση, επιφάνεια Laplacian, κ.λπ.).

Εξαγωγή χαρακτηριστικών: Μόλις προ επεξεργαστούν σήματα, μπορούν να εξαχθούν χαρακτηριστικά. Αυτά τα χαρακτηριστικά αποτελούνται από μερικές τιμές που περιγράφουν τις σχετικές πληροφορίες που είναι ενσωματωμένες στα σήματα (Bashashati et al., 2007), όπως η ισχύς των σημάτων σε συγκεκριμένες ζώνες

συχνοτήτων (Pfurtscheller & Neuper, 2001). Αυτά τα χαρακτηριστικά συγκεντρώνονται έπειτα σε ένα διάνυσμα που ονομάζεται «διάνυσμα χαρακτηριστικών». Παραδείγματα δυνατοτήτων που διατίθενται στο OpenViBE περιλαμβάνουν χαρακτηριστικά ισχύος ζώνης ή πυκνότητες φάσματος ισχύος.

Ταξινόμηση: Το διάνυσμα χαρακτηριστικών τροφοδοτείται σε έναν αλγόριθμο γνωστό ως ταξινομητής.

Ένας ταξινομητής εκχωρεί μια κλάση σε κάθε διάνυσμα χαρακτηριστικών, με αυτήν την τάξη να είναι αναγνωριστικό του εγκεφαλικού σήματος που έχει αναγνωριστεί. Γενικά, ο ταξινομητής εκπαιδεύεται εκ των προτέρων χρησιμοποιώντας ένα σύνολο διανυσμάτων χαρακτηριστικών από κάθε τάξη. Ένα παράδειγμα ταξινομητή που χρησιμοποιείται για το BCI θα ήταν η Γραμμική Ανάλυση Διακριτών (Lotte, Congedo, L'ecuyer, Lamarche, & Arnaldi, 2007). Πρέπει να σημειωθεί ότι, λόγω της μεγάλης μεταβλητότητας και του θορύβου των σημάτων EEG, πολύ γρήγορα επιτυγχάνονται ποσοστά ταξινόμησης 100%, ακόμη και για ένα BCI που χρησιμοποιεί δύο νοητικές καταστάσεις. Το OpenViBE προτείνει διάφορους ταξινομητές όπως η Linear Discriminant Analysis (Lotte, Congedo, et al., 2007) ή Fuzzy Inference Systems (Lotte, L'ecuyer, Lamarche, & Arnaldi, 2007).

Μετάφραση σε μια εντολή: Μόλις προσδιοριστεί η κλάση του σήματος, μπορεί να συσχετιστεί με μια εντολή που αποστέλλεται σε έναν υπολογιστή προκειμένου να ελέγξει, για παράδειγμα, ένα ρομπότ (Mill'an, 2008) ή μια πρόσθεση (Wolpaw et al., 2002). Ο αριθμός πιθανών εντολών στα τρέχοντα συστήματα BCI που βασίζονται σε EEG κυμαίνεται συνήθως μεταξύ 1 και 4.

Ανατροφοδότηση: Τέλος, θα πρέπει να παρέχονται σχόλια στον χρήστη, ώστε να μπορεί να καθορίσει εάν έκανε σωστά το εγκεφαλικό σήμα. Αυτό είναι ένα σημαντικό βήμα καθώς βοηθά τον χρήστη να ελέγχει τη δραστηριότητα του εγκεφάλου του (Lotte, Renard, & L'ecuyer, 2008) (Neuper, Scherer, Wriessnegger, & Pfurtscheller, 2009). Τα σχόλια μπορεί να είναι απλά οπτικά ή ακουστικά στοιχεία, π.χ. μετρητές. Για το σκοπό αυτό, το λογισμικό μας προτείνει κλασσικές ενότητες ακατέργαστου σήματος, φασμάτων, χρόνου / συχνότητας. Εναλλακτικά, πιο προηγμένα σχόλια μπορούν να παρέχονται όπως η τροποποίηση ενός εικονικού περιβάλλοντος (Leeb et al., 2007) στο οποίο το OpenViBE στέλνει εντολές.

3.2 Το Λογισμικό Weka για Εξόρυξη Δεδομένων

Το Waikato Environment for Knowledge Analysis (Weka), που αναπτύχθηκε στο Πανεπιστήμιο του Waikato της Νέας Ζηλανδίας, είναι ελεύθερο λογισμικό με άδεια χρήσης GNU General Public License και το συνοδευτικό λογισμικό στο βιβλίο Data Mining: Practical Machine Learning Tools and Techniques (Witten, Ian H.;2011).

Το Weka περιέχει μια συλλογή εργαλείων οπτικοποίησης και αλγορίθμων για ανάλυση δεδομένων και προγνωστική μοντελοποίηση, μαζί με γραφικές διεπαφές χρήστη για εύκολη πρόσβαση σε αυτές τις λειτουργίες (Witten, Ian H.; 2011). Η αρχική έκδοση του Weka εκτός Java ήταν ένα αλγόριθμο μοντελοποίησης Tcl / Tk (κυρίως τρίτων) που εφαρμόστηκε σε άλλες γλώσσες προγραμματισμού, καθώς και βοηθητικά προγράμματα προεπεξεργασίας δεδομένων στο C και ένα σύστημα βασισμένο στο Makefile για την εκτέλεση πειραμάτων μηχανικής μάθησης. Αυτή η αρχική έκδοση σχεδιάστηκε κυρίως ως εργαλείο για την ανάλυση δεδομένων από γεωργικούς τομείς (Holmes, Geoffrey 1994, Garner, Stephen R 1995), αλλά η πιο πρόσφατη πλήρως βασισμένη σε Java έκδοση (Weka 3), για την οποία ξεκίνησε η ανάπτυξη το 1997, χρησιμοποιείται τώρα σε πολλές διαφορετικές εφαρμογές τομείς, ιδίως για εκπαιδευτικούς σκοπούς και έρευνα. Τα πλεονεκτήματα του Weka περιλαμβάνουν:

- ✚ Δωρεάν διαθεσιμότητα βάσει της άδειας GNU General Public.
- ✚ Φορητότητα, δεδομένου ότι εφαρμόζεται πλήρως στη γλώσσα προγραμματισμού Java και λειτουργεί έτσι σε σχεδόν οποιαδήποτε σύγχρονη υπολογιστική πλατφόρμα.
- ✚ Μια ολοκληρωμένη συλλογή τεχνικών προεπεξεργασίας και μοντελοποίησης δεδομένων.
- ✚ Ευκολία χρήσης λόγω των γραφικών διεπαφών χρήστη.

Η Weka υποστηρίζει πολλές τυπικές εργασίες εξόρυξης δεδομένων, πιο συγκεκριμένα, προεπεξεργασία δεδομένων, ομαδοποίηση, ταξινόμηση, παλινδρόμηση, οπτικοποίηση και επιλογή χαρακτηριστικών. Όλες οι τεχνικές του Weka βασίζονται στην υπόθεση ότι τα δεδομένα είναι διαθέσιμα ως ένα ενιαίο αρχείο ή σχέση, όπου κάθε σημείο δεδομένων περιγράφεται από έναν καθορισμένο αριθμό χαρακτηριστικών (συνήθως, αριθμητικά ή ονομαστικά χαρακτηριστικά, αλλά υποστηρίζονται και ορισμένοι άλλοι τύποι χαρακτηριστικών). Η Weka παρέχει πρόσβαση σε βάσεις δεδομένων SQL

χρησιμοποιώντας Java Database Connectivity και μπορεί να επεξεργαστεί το αποτέλεσμα που επιστρέφεται από ένα ερώτημα βάσης δεδομένων. Το Weka παρέχει πρόσβαση στη βαθιά μάθηση με το Deeplearning4j (Reutemann, Peter; 2004). Δεν είναι ικανό για πολλαπλές σχέσεις εξόρυξης δεδομένων, αλλά υπάρχει ξεχωριστό λογισμικό για τη μετατροπή μιας συλλογής συνδεδεμένων πινάκων βάσεων δεδομένων σε έναν ενιαίο πίνακα που είναι κατάλληλος για επεξεργασία χρησιμοποιώντας το Weka (Witten, Ian H 1999). Ένας άλλος σημαντικός τομέας που επί του παρόντος δεν καλύπτεται από τους αλγόριθμους που περιλαμβάνονται στη διανομή Weka είναι η μοντελοποίηση ακολουθιών.

ΚΕΦΑΛΑΙΟ 4 ΜΕΘΟΔΟΛΟΓΙΑ

4.1 Μηχανική Μάθηση και data mining

Η μηχανική μάθηση ως επιστημονική πειθαρχία εξακολουθεί να αναδύεται και συνεπώς υφίσταται συνεχείς αλλαγές. Ενώ πολλές από τις μεθόδους και τους αλγορίθμους που χρησιμοποιήθηκαν είναι γνωστές εδώ και δεκαετίες, τα τελευταία χρόνια, οι νέες προσεγγίσεις ωρίμασαν σε βαθμό που είναι έγκυρο να θεωρηθεί η μηχανική εκμάθηση ένα νέο και ακόμα νεογέννητο πεδίο, παρά την ήδη ολοκληρωμένη ανάπτυξή του για μια σημαντική περίοδο χρόνου. Ως εκ τούτου, αυτό που συνιστά τη μηχανική εκμάθηση ακριβώς (σε αντίθεση με, π.χ., περιγραφικά στατιστικά στοιχεία) παραμένει μόνο ασαφές. Με τις μεθοδολογίες που βασίζονται σε δεδομένα να ενσωματώνονται σε τομείς όπως η επιστήμη των υλικών, νέες παραλλαγές και προσαρμοσμένες μέθοδοι μηχανικής μάθησης έχουν επινοηθεί ή βρίσκονται στη διαδικασία προσαρμογής στις προκλήσεις και τα προφίλ δεδομένων που είναι μοναδικά για την επιστήμη των υλικών. Αυτή η μεθοδολογική εξειδίκευση τομέα δεν πρέπει να εκληφθεί ότι αποκλείει τη σημασία των βασικών μεθόδων της μηχανικής μάθησης, όπως τα τεχνητά νευρικά δίκτυα, τα οποία θεωρητικά είναι για όλες τις χρήσεις και προσαρμόζονται στην προσέγγιση (μάθηση) οποιασδήποτε λειτουργίας που είναι εγγενής στα δεδομένα [Universal Θεώρημα προσέγγισης (Hornik, 1991)]. Ωστόσο, καθώς οι προσεγγίσεις από την επιστήμη των δεδομένων αυξάνονται και συγχωνεύονται με τις παραδοσιακές ερευνητικές διαδικασίες της επιστήμης των υλικών, οι μέθοδοι που αναφέρονται σε αυτό το κεφάλαιο δεν μπορούν να ισχυριστούν ότι είναι μια εξαντλητική απαρίθμηση μεθόδων μηχανικής μάθησης που είναι βιώσιμες για τη (συνέχεια) μηχανική υλικών, ως συνεχείς αλλαγές αναμένονται μέσα στα επόμενα χρόνια.

Σε αυτό το πλαίσιο, η πιο συχνά συναντούμενη τάξη μηχανικής μάθησης (στο βαθμό που μερικές φορές χρησιμοποιείται εναλλακτικά με τον ίδιο τον όρο μηχανική εκμάθηση) είναι η τάξη των τεχνητών νευρικών δικτύων (ANN). Προερχόμενος από μια απλή πρόδρομη διατύπωση που χρονολογείται από το 1958, το perceptron (Rosenblatt, 1958), τα ANN έχουν κερδίσει τη δημοτικότητά τους καθώς η αυξανόμενη

υπολογιστική ισχύς και η διαθεσιμότητα δεδομένων ανακουφίζουν τα δύο σημεία συμφόρησης που προηγουμένως μείωσαν τη χρήση τους. Το ίδιο το perceptron σχεδιάστηκε ως ένα απλό νευρωνικό δίκτυο ενός επιπέδου και χρησιμοποιήθηκε ως γραμμικός ταξινομητής.

Στην απλούστερη σύγχρονη μορφή τους, τα νευρικά δίκτυα τροφοδότησης προς τα εμπρός (FFNNs) (Haykin, 1998; Russell et al., 2016) είναι πολυεπίεδα περιπετρώνια, δηλαδή στρώματα κορυφών (νευρώνες) στα οποία κάθε νευρώνας υπολογίζει μια έξοδο με βάση τις εισόδους από το προηγούμενο στρώμα. Το σήμα διασχίζει το δίκτυο με μονοκατευθυντικό τρόπο, μετατρέποντας σταδιακά το σήμα εισόδου σε σήμα εξόδου καθώς διεισδύει μέσω του δικτύου, οπότε ονομάζεται feedforward. Τέτοια δίκτυα εκπαιδεύονται συνήθως χρησιμοποιώντας μια προσέγγιση ελαχιστοποίησης σφαλμάτων διαβάθμισης πίσω. Το σφάλμα προσδιορίζεται συγκρίνοντας την τρέχουσα έξοδο δικτύου με τη σωστή έξοδο (η οποία είναι διαθέσιμη σε ένα εποπτευόμενο σενάριο μάθησης). Στη συνέχεια, η ατομική συμπεριφορά νευρώνων αλλάζει κατά τη διάρκεια της διαδικασίας εκπαίδευσης / εκμάθησης αλλάζοντας τα βάρη σύνδεσης που σχετίζονται με κάθε άκρη στην τοπολογία του δικτύου. Υπάρχουν πολλές διαφορετικές προσεγγίσεις για την αντιμετώπιση των διαφόρων δυσκολιών που είναι συνήθως εγγενείς στα νευρικά δίκτυα, για παράδειγμα, υπερπροσφορά, τοπικά ελάχιστα, προσδιορισμός του βέλτιστου αριθμού στρωμάτων και νευρώνων ανά στρώματα, επιλογή λειτουργίας ενεργοποίησης και ανθρώπινη ερμηνεία του δικτύου, μεταξύ άλλων.

Μόλις στοιβαχτούν αρκετά στρώματα νευρώνων, τα οποία μπορούν να αναφέρονται ως το βάθος του υποκείμενου ANN, η συμπεριφορά του δικτύου εισέρχεται στο καθεστώς της βαθιάς μάθησης (DL). Με αυξημένο αριθμό στρωμάτων, οι νέες δυσκολίες έρχονται στο επίκεντρο, όπως το πρόβλημα της διαγραφής βαθμίδωσης (Hochreiter, 1998) ή ο χρόνος υπολογισμού, μεταξύ άλλων, που δεν λείπουν από τις απλούστερες αρχιτεκτονικές αλλά επιδεινώνονται στην περίπτωση του DL. Οι αρχιτεκτονικές DL είναι, ως επί το πλείστον, αγνωστικές σχετικά με τον τύπο του ANN, δηλαδή, οποιοδήποτε είδος ANN μπορεί να αποτελέσει τη βάση για μια αρχιτεκτονική DL. Δύο υπότυποι των αρχιτεκτονικών ANN έχουν κερδίσει τη δημοτικότητα τους ιδιαίτερα τις τελευταίες δύο δεκαετίες, στο βαθμό που για περίπλοκα προβλήματα, συνήθως ένα από τα δύο αντιμετωπίζεται τουλάχιστον ως

συστατικό της συνολικής αρχιτεκτονικής ANN, ή ως προ-/ μετά την επεξεργασία βήμα του αγωγού μάθησης. Αυτά είναι συνελκτικά νευρικά δίκτυα και μακροπρόθεσμα δίκτυα μνήμης.

Τα κυριαρχικά νευρωνικά δίκτυα (CNNs) έχουν ως επί το πλείστον διαμορφωθεί ως παραλλαγή των FFNNs και πρωτοπορήθηκαν το 1980 (Fukushima, 1980) και στη συνέχεια αναδιαμορφώθηκαν στη σύγχρονη μορφή τους το 1999 (LeCun et al., 1999). Ταιριάζουν ιδιαίτερα καλά για την αναγνώριση εικόνας, δηλαδή την αναγνώριση προτύπων σε οπτικά δεδομένα (Schmidhuber, 2015; Russell et al., 2016). Συνήθως, ένα συνελκτικό στρώμα μετατοπίζεται κατά μήκος των δεδομένων που μοιάζουν με ένα φίλτρο / ανιχνευτή σε αλγόριθμους όρασης υπολογιστή, απαιτώντας μόνο λίγες παραμέτρους λόγω του συνεχόμενου στρώματος που επιτρέπει αποτελεσματική αναπαραγωγή βάρους καθώς το «φίλτρο» αναπαράγεται σε όλο το οπτικό πεδίο. Τα επίπεδα συγκέντρωσης και ομαλοποίησης επιτρέπουν τη σταδιακή απλοποίηση δεδομένων και για μεταβλητά μεγέθη χαρακτηριστικών, αντίστοιχα. Παρόλο που η καταλληλότητα των CNN για την επιστήμη υλικών ενδέχεται να μην είναι άμεσα εμφανής, υπάρχουν παραδείγματα άμεσων εφαρμογών όπως η αναγνώριση υφής υλικών (Cang and Ren, 2016; Lubbers et al., 2017; Cecen et al., 2018), έμμεσα παραδείγματα εφαρμογών στα οποία μπορεί να ερμηνευθούν δεδομένα μη οπτικού υλικού (Schwarzer et al., 2019).

Η μακροχρόνια βραχυπρόθεσμη μνήμη (LSTM) (Hochreiter and Schmidhuber, 1997; Russell et al., 2016) μπορεί να συνθέσει ANNs επειδή προσφέρουν εξειδικευμένους νευρώνες / μονάδες μνήμης που ασχολούνται κυρίως με το πρόβλημα βαθμίδωσης που εξαφανίζεται (ή αντιστρόφως το πρόβλημα της έκρηξης που εκρήγνυται), που συχνά οδηγούν σε μη βέλτιστα τοπικά ελάχιστα, ειδικά καθώς αυξάνεται ο αριθμός των νευρωνικών στρωμάτων. Καθώς οι αρχιτεκτονικές βαθιάς μάθησης γίνονται όλο και πιο δημοφιλείς, τα LSTM ή οι παραλλαγές τους έχουν κερδίσει τη δημοτικότητά τους σε κλειδαριά με το DL ως τρόπο παράκαμψης τέτοιων τοπικών παγίδων. Εν ολίγοις, η «αποθήκευση» σημαντικών σημείων δεδομένων με την πάροδο του χρόνου από το να πνιγεί και η διανομή του σήματος διόρθωσης σφαλμάτων τους για μεγαλύτερες περιόδους επιτρέπει καλύτερη αποθήκευση πληροφοριών σχετικά με σημαντικά συμβάντα.

Για δυναμικά προβλήματα στα οποία ένα σημείο δεδομένων συχνά εκτίθεται σε μια χρονική εξέλιξη της κατάστασης των υλικών και περιλαμβάνει μια σειρά ενεργειών, προτιμώνται άλλες προσεγγίσεις εκτός από τη στατική περίπτωση. Ενώ πολλά από αυτά δεν έχουν μεταναστεύσει στην επιστήμη των υλικών σε σημαντικό βαθμό, ωστόσο, μπορεί να αναμένεται ότι οι προσαρμοσμένες μέθοδοι που είναι κατάλληλες για δυναμικά προβλήματα θα αποκτήσουν αυξανόμενη σημασία, τόσο καθώς αυξάνεται η πολυπλοκότητα των λειτουργιών στόχου που πρέπει να μάθει, όσο και με την ενσωμάτωση των εξαρτήσεων κατάστασης (συνήθως ως συνάρτηση του χρόνου). Δύο από τις κύριες προσεγγίσεις σε τέτοια προβλήματα είναι η ενίσχυση της μάθησης, η οποία αποτελείται από διαφορετικές μεθόδους όπως η Q-learning (Watkins and Dayan, 1992) και τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) (Lipton, 2015; Russell et al., 2016), που επιτρέπουν κατευθυνόμενους κύκλους εντός της τοπολογίας ANN, και έτσι για σήματα να ταλαντεύονται και να αλληλεπικαλύπτονται με τον υπολογισμό των επόμενων δειγμάτων. Ως αποτέλεσμα, τα δεδομένα διαβιβάζονται επιλεκτικά σε βήματα ακολουθίας.

Αξίζει να σημειωθεί είναι τυχαιοποιημένα νευρικά δίκτυα, τα οποία προσθέτουν τυχαίες αιχμές διέγερσης / αναστολής σε μεμονωμένους νευρώνες (Gallicchio et al., 2017) χωρίς σταθερές εσωτερικές καταστάσεις (Maass et al., 2002) και νευρικά δίκτυα ακτινικής βάσης (Ott, 1996), τα οποία είναι τυπικά ρηχά FFNN που χρησιμοποιούν μεμονωμένες λειτουργίες ακτινικής βάσης ειδικών νευρώνων για να αθροίσουν τις εισόδους νευρώνων και έτσι να επιτρέψουν καλύτερη εξειδίκευση μεμονωμένων νευρώνων.

Μία αξιοσημείωτη εναλλακτική λύση στα νευρωνικά δίκτυα είναι το Support Vector Machines (SVMs), που εισήχθησαν από τους Cortes και Vapnik (1995). Οι SVMs αποκλίνουν από τα ANNs χωρίς να χαρτογραφούν ούτε σε συνεχή (πρόβλημα παλινδρόμησης) ούτε σε διακριτή (πρόβλημα απόφασης) έξοδο, αλλά μάλλον διαχωρίζοντας μοτίβα είτε μέσω υπερπλάνων στη γραμμικά διαχωρισμένη θήκη είτε σε λειτουργία Kernel στη μη γραμμικά διαχωρισμένη περίπτωση. Αυτός ο χάρτης μετασχηματισμού πυρήνα («τέχνασμα») υποστηρίζει διανύσματα σε έναν μετασχηματισμένο χώρο χαρακτηριστικών στον οποίο είναι δυνατή η διαχωριστικότητα (χρησιμοποιώντας, π.χ., ένα μέγιστο περιθώριο) Ενώ, τα ANN

χρησιμοποιούν συνήθως πολλά στρώματα, συχνά αποτελούμενα από απλούς νευρώνες, τα SVM μπορούν να ερμηνευτούν ως ένας εξειδικευμένος μονός νευρικός κόμβος.

4.2 Σύνολο Δεδομένων

Η συλλογή των δεδομένων θα γίνει με την ηχογράφηση κόπου οι συμμετέχοντες αποτελούνται από τέσσερα άτομα χ γυναίκες και ψ άντρες ηλικίας από χ ως ψ .

Στους συμμετέχοντες θα δοθεί ένα κείμενο να το διαβάσουν και έχει τηρηθεί το απόρρητο των προσωπικών δεδομένων. Οι συμμετέχοντες έλαβαν μέρος εθελοντικά και τους επισημάνθηκε ότι τα δεδομένα θα χρησιμοποιηθούν μόνο για την έρευνα.

4.4 Μέθοδοι Ταξινόμησης (Περιγραφή Αλγορίθμων)

4.4.1 Linear Discriminant analysis (LDA)

Η ανάλυση γραμμικής διάκρισης (LDA), η κανονική ανάλυση διακρίσεων (NDA) ή η ανάλυση λειτουργίας διακρίσεων είναι μια γενίκευση της γραμμικής διάκρισης του Fisher, μια μέθοδος που χρησιμοποιείται στις στατιστικές, την αναγνώριση προτύπων και άλλα πεδία, για την εύρεση ενός γραμμικού συνδυασμού χαρακτηριστικών που χαρακτηρίζει ή διαχωρίζει δύο ή περισσότερες κατηγορίες αντικειμένων ή συμβάντων. Ο προκύπτων συνδυασμός μπορεί να χρησιμοποιηθεί ως γραμμικός ταξινομητής ή, συνηθέστερα, για μείωση διαστάσεων πριν από την μετέπειτα ταξινόμηση.

Το LDA σχετίζεται στενά με την ανάλυση διακύμανσης (ANOVA) και την ανάλυση παλινδρόμησης, η οποία προσπαθεί επίσης να εκφράσει μια εξαρτημένη μεταβλητή ως γραμμικό συνδυασμό άλλων χαρακτηριστικών ή μετρήσεων (Fisher, R. A 1936, McLachlan, G. J, 2004). Ωστόσο, η ANOVA χρησιμοποιεί κατηγορηματικές ανεξάρτητες μεταβλητές και μια συνεχή εξαρτώμενη μεταβλητή, ενώ η διακριτική ανάλυση έχει συνεχείς ανεξάρτητες μεταβλητές και μια κατηγορηματική εξαρτώμενη μεταβλητή (δηλαδή την ετικέτα κλάσης) (Analyzing Quantitative Data). Η λογιστική παλινδρόμηση και η παλινδρόμηση probit μοιάζουν περισσότερο με το LDA από ό, τι είναι το ANOVA, καθώς εξηγούν επίσης μια κατηγορηματική μεταβλητή από τις τιμές των συνεχών ανεξάρτητων μεταβλητών. Αυτές οι άλλες μέθοδοι είναι προτιμότερες σε

εφαρμογές όπου δεν είναι λογικό να υποθέσουμε ότι οι ανεξάρτητες μεταβλητές είναι κανονικά κατανομημένες, κάτι που αποτελεί θεμελιώδη υπόθεση της μεθόδου LDA.

Το LDA σχετίζεται επίσης στενά με την ανάλυση βασικών συστατικών (PCA) και την ανάλυση παραγόντων στο ότι και οι δύο αναζητούν γραμμικούς συνδυασμούς μεταβλητών που εξηγούν καλύτερα τα δεδομένα (Martinez, A. M, 2001). Το LDA προσπαθεί ρητά να μοντελοποιήσει τη διαφορά μεταξύ των κατηγοριών δεδομένων. Το PCA, σε αντίθεση, δεν λαμβάνει υπόψη καμία διαφορά στην τάξη, και η ανάλυση παραγόντων δημιουργεί τους συνδυασμούς δυνατοτήτων με βάση διαφορές παρά ομοιότητες. Η ανάλυση διακρίσεων είναι επίσης διαφορετική από την ανάλυση παραγόντων στο ότι δεν είναι τεχνική αλληλεξάρτησης: πρέπει να γίνει διάκριση μεταξύ ανεξάρτητων μεταβλητών και εξαρτημένων μεταβλητών (που ονομάζονται επίσης μεταβλητές κριτηρίου).

Το LDA λειτουργεί όταν οι μετρήσεις που γίνονται σε ανεξάρτητες μεταβλητές για κάθε παρατήρηση είναι συνεχείς ποσότητες. Όταν ασχολούμαστε με κατηγορηματικές ανεξάρτητες μεταβλητές, η ισοδύναμη τεχνική είναι διακριτική ανάλυση αλληλογραφίας (Abdi, H.2007, Perriere, G.;2003).

Η ανάλυση διακρίσεων χρησιμοποιείται όταν οι ομάδες είναι γνωστές εκ των προτέρων (σε αντίθεση με την ανάλυση συστάδων). Κάθε περίπτωση πρέπει να έχει βαθμολογία σε ένα ή περισσότερα ποσοτικά μέτρα πρόβλεψης και βαθμολογία σε ομαδική μέτρηση (BÖKEOĞLU ÇOKLUK, Ö, 2008). Με απλούς όρους, η ανάλυση διακριτικής λειτουργίας είναι η ταξινόμηση - η πράξη της διανομής των πραγμάτων σε ομάδες, τάξεις ή κατηγορίες του ίδιου τύπου.

4.4.2 Νευρωνικά δίκτυα (Perceptron)

Ένα νευρωνικό δίκτυο είναι ένα δίκτυο ή κύκλωμα νευρώνων, ή με μια σύγχρονη έννοια, ένα τεχνητό νευρωνικό δίκτυο, αποτελούμενο από τεχνητούς νευρώνες ή κόμβους (Hopfield, J. J.1982). Έτσι, ένα νευρικό δίκτυο είναι είτε ένα βιολογικό νευρωνικό δίκτυο, που αποτελείται από πραγματικούς βιολογικούς νευρώνες, είτε ένα τεχνητό νευρωνικό δίκτυο, για την επίλυση προβλημάτων τεχνητής νοημοσύνης (AI). Οι συνδέσεις του βιολογικού νευρώνα διαμορφώνονται ως βάρη. Ένα θετικό βάρος

αντανακλά μια διεγερτική σύνδεση, ενώ οι αρνητικές τιμές σημαίνουν ανασταλτικές συνδέσεις. Όλες οι είσοδοι τροποποιούνται κατά βάρος και αθροίζονται. Αυτή η δραστηριότητα αναφέρεται ως γραμμικός συνδυασμός. Τέλος, μια λειτουργία ενεργοποίησης ελέγχει το πλάτος της εξόδου. Για παράδειγμα, ένα αποδεκτό εύρος εξόδου είναι συνήθως μεταξύ 0 και 1, ή θα μπορούσε να είναι -1 και 1 .

Αυτά τα τεχνητά δίκτυα μπορούν να χρησιμοποιηθούν για προγνωστική μοντελοποίηση, προσαρμοστικό έλεγχο και εφαρμογές όπου μπορούν να εκπαιδευτούν μέσω ενός συνόλου δεδομένων. Η αυτομάθηση που προκύπτει από την εμπειρία μπορεί να συμβεί μέσα σε δίκτυα, τα οποία μπορούν να εξαγάγουν συμπεράσματα από ένα πολύπλοκο και φαινομενικά άσχετο σύνολο πληροφοριών.

Ένα βιολογικό νευρωνικό δίκτυο αποτελείται από ομάδες νευρώνων που συνδέονται χημικά ή λειτουργικά. Ένας μεμονωμένος νευρώνας μπορεί να συνδεθεί με πολλούς άλλους νευρώνες και ο συνολικός αριθμός νευρώνων και συνδέσεων σε ένα δίκτυο μπορεί να είναι εκτεταμένος. Οι συνδέσεις, που ονομάζονται συνάψεις, συνήθως σχηματίζονται από άξονες σε δενδρίτες, αν και δενδενδριδικές συνάψεις (Cuntz, Hermann, 2010) και άλλες συνδέσεις είναι δυνατές. Εκτός από την ηλεκτρική σηματοδότηση, υπάρχουν και άλλες μορφές σηματοδότησης που προκύπτουν από τη διάχυση νευροδιαβιβαστών.

Η τεχνητή νοημοσύνη, η γνωστική μοντελοποίηση και τα νευρωνικά δίκτυα είναι παραδείγματα επεξεργασίας πληροφοριών εμπνευσμένα από τον τρόπο με τον οποίο τα βιολογικά νευρικά συστήματα επεξεργάζονται δεδομένα. Η τεχνητή νοημοσύνη και η γνωστική μοντελοποίηση προσπαθούν να προσομοιώσουν ορισμένες ιδιότητες των βιολογικών νευρωνικών δικτύων. Στον τομέα της τεχνητής νοημοσύνης, τα τεχνητά νευρικά δίκτυα έχουν εφαρμοστεί με επιτυχία στην αναγνώριση ομιλίας, στην ανάλυση εικόνων και στον προσαρμοστικό έλεγχο, προκειμένου να κατασκευαστούν παράγοντες λογισμικού (σε υπολογιστές και βιντεοπαιχνίδια) ή αυτόνομα ρομπότ.

Ιστορικά, οι ψηφιακοί υπολογιστές εξελίχθηκαν από το μοντέλο von Neumann και λειτουργούν μέσω της εκτέλεσης ρητών οδηγιών μέσω της πρόσβασης στη μνήμη από έναν αριθμό επεξεργαστών. Από την άλλη πλευρά, η προέλευση των νευρωνικών δικτύων βασίζεται σε προσπάθειες μοντελοποίησης της επεξεργασίας πληροφοριών σε

βιολογικά συστήματα. Σε αντίθεση με το μοντέλο von Neumann, ο υπολογιστής νευρικού δικτύου δεν διαχωρίζει τη μνήμη και την επεξεργασία (Yang, J. J.;2008).

Η θεωρία του νευρικού δικτύου έχει χρησιμεύσει τόσο για τον καλύτερο προσδιορισμό του τρόπου λειτουργίας των νευρώνων στον εγκέφαλο όσο και για τη δημιουργία βάσεων για προσπάθειες δημιουργίας τεχνητής νοημοσύνης.

4.4.3 Μηχανές Διανυσμάτων υποστήριξης (*Support Vector Machines - SVM*)

Στη μηχανική μάθηση, οι μηχανές υποστήριξης-φορέα (SVMs, καθώς και τα δίκτυα φορέα υποστήριξης είναι εποπτευόμενα μοντέλα μάθησης με συναφείς αλγόριθμους μάθησης που αναλύουν δεδομένα που χρησιμοποιούνται για την ταξινόμηση και την ανάλυση παλινδρόμησης (Cortes, Corinna 1995). Αναπτύχθηκε στα AT&T Bell Laboratories από τον Vapnik με συναδέλφους (Boser et al., 1992, Guyon et al., 1993, Vapnik et al., 1997), παρουσιάζει μια από τις πιο ισχυρές μεθόδους πρόβλεψης, με βάση το πλαίσιο στατιστικής μάθησης ή το VC θεωρία που προτάθηκε από τους Vapnik και Chervonenkis (1974) και Vapnik (1982, 1995). Λαμβάνοντας υπόψη ένα σύνολο παραδειγμάτων εκπαίδευσης, το καθένα επισημαίνεται ότι ανήκει σε μία ή την άλλη από δύο κατηγορίες, ένας αλγόριθμος εκπαίδευσης SVM δημιουργεί ένα μοντέλο που εκχωρεί νέα παραδείγματα σε μια κατηγορία ή την άλλη, καθιστώντας τον μη πιθανό δυαδικό γραμμικό ταξινομητή (αν και μέθοδοι όπως η κλιμάκωση Platt υπάρχουν για τη χρήση SVM σε μια πιθανολογική ρύθμιση ταξινόμησης). Ένα μοντέλο SVM είναι μια αναπαράσταση των παραδειγμάτων ως σημεία στο διάστημα, χαρτογραφημένα έτσι ώστε τα παραδείγματα των ξεχωριστών κατηγοριών να διαιρούνται με ένα σαφές κενό που είναι όσο το δυνατόν ευρύτερο. Στη συνέχεια, νέα παραδείγματα χαρτογραφούνται στον ίδιο χώρο και προβλέπεται να ανήκουν σε μια κατηγορία με βάση την πλευρά του κενού στην οποία πέφτουν.

Εκτός από την εκτέλεση γραμμικής ταξινόμησης, τα SVM μπορούν να εκτελέσουν αποτελεσματικά μια μη γραμμική ταξινόμηση χρησιμοποιώντας αυτό που ονομάζεται κόλπο πυρήνα, χαρτογραφώντας σιωπηρά τις εισόδους τους σε χώρους διαστάσεων υψηλής διάστασης.

Όταν τα δεδομένα δεν επισημαίνονται, η εποπτευόμενη μάθηση δεν είναι δυνατή και απαιτείται μια μη εποπτευόμενη μαθησιακή προσέγγιση, η οποία επιχειρεί να βρει φυσική ομαδοποίηση των δεδομένων σε ομάδες και στη συνέχεια να χαρτογραφήσει νέα δεδομένα σε αυτές τις σχηματισμένες ομάδες. Ο αλγόριθμος ομαδοποίησης υποστήριξης-φορέα (Ben-Hur, 2001), που δημιουργήθηκε από τους Hava Siegelmann και Vladimir Varnik, εφαρμόζει τα στατιστικά στοιχεία διανυσμάτων υποστήριξης, που αναπτύχθηκαν στον αλγόριθμο μηχανών φορέα υποστήριξης, για την κατηγοριοποίηση δεδομένων χωρίς ετικέτα και είναι ένας από τους ευρύτερα χρησιμοποιούμενους αλγόριθμους συμπλέγματος σε βιομηχανικούς εφαρμογές.

4.4.4 Δένδρα Απόφασης – (Περιγραφή Αλγορίθμου C4.5)

Το C4.5 είναι ένας αλγόριθμος που χρησιμοποιείται για τη δημιουργία ενός δέντρου αποφάσεων που αναπτύχθηκε από τον Ross Quinlan (Quinlan, J. R 1993). Το C4.5 είναι μια επέκταση του προηγούμενου αλγορίθμου ID3 της Quinlan Τα δέντρα αποφάσεων που δημιουργούνται από το C4.5 μπορούν να χρησιμοποιηθούν για ταξινόμηση και για αυτόν τον λόγο, το C4.5 αναφέρεται συχνά ως στατιστικός ταξινομητής. Το 2011, οι συγγραφείς του λογισμικού μηχανικής μάθησης Weka περιέγραψαν τον αλγόριθμο C4.5 ως ένα ορόσημο πρόγραμμα αποφάσεων ορόσημο που είναι πιθανότατα το εργαστήριο μηχανικής εκμάθησης που χρησιμοποιείται ευρέως στην πράξη μέχρι σήμερα (Ian H, 2011).

Έγινε αρκετά δημοφιλές μετά την κατάταξη # 1 στους κορυφαίους 10 αλγόριθμους στην εξέχουσα εργασία Mining Data που δημοσιεύθηκε από το Springer LNCS το 2008 (S.B. Kotsiantis 2007).

Το C4.5 δημιουργεί δέντρα αποφάσεων από ένα σύνολο δεδομένων εκπαίδευσης με τον ίδιο τρόπο όπως το ID3, χρησιμοποιώντας την έννοια της εντροπίας πληροφοριών. Τα δεδομένα εκπαίδευσης είναι ένα σύνολο $S = \{s_1, s_2, \dots\}$ ήδη ταξινομημένων δειγμάτων. Κάθε δείγμα s_i αποτελείται από ένα πολυδιάστατο διάνυσμα $(x_{1,i}, x_{2,i}, \dots, x_{p,i})$, όπου το x_j αντιπροσωπεύει τιμές χαρακτηριστικών ή χαρακτηριστικά του δείγματος, καθώς και την κλάση στην οποία πέφτει το s_i .

Σε κάθε κόμβο του δέντρου, το C4.5 επιλέγει το χαρακτηριστικό των δεδομένων που χωρίζει αποτελεσματικότερα το σύνολο δειγμάτων του σε υποσύνολα εμπλουτισμένα σε μία κατηγορία ή στην άλλη. Το κριτήριο διαχωρισμού είναι το ομαλοποιημένο κέρδος πληροφοριών (διαφορά εντροπίας). Το χαρακτηριστικό με το υψηλότερο κανονικοποιημένο κέρδος πληροφοριών επιλέγεται για τη λήψη της απόφασης. Ο αλγόριθμος C4.5 επαναλαμβάνεται στη συνέχεια στις κατατμημένες δευτερεύουσες λίστες.

Αυτός ο αλγόριθμος έχει μερικές βασικές περιπτώσεις.

Όλα τα δείγματα στη λίστα ανήκουν στην ίδια τάξη. Όταν συμβαίνει αυτό, δημιουργεί απλά έναν κόμβο φύλλων για το δέντρο αποφάσεων που λέει ότι επιλέγει αυτήν την τάξη.

Καμία από τις δυνατότητες δεν παρέχει κέρδος πληροφοριών. Σε αυτήν την περίπτωση, το C4.5 δημιουργεί έναν κόμβο απόφασης ψηλότερα στο δέντρο χρησιμοποιώντας την αναμενόμενη τιμή της κλάσης.

Παρουσιάστηκε παρουσία προηγούμενης αόρατης τάξης. Και πάλι, το C4.5 δημιουργεί έναν κόμβο απόφασης ψηλότερα στο δέντρο χρησιμοποιώντας την αναμενόμενη τιμή.

BIBΛΙΟΓΡΑΦΙΑ

- Hodgson, Jay (2010). *Understanding Records*, p.1. ISBN 978-1-4411-5607-5.
- Atti, Andreas Spanias, Ted Painter, Venkatraman (2006). *Audio signal processing and coding* ([Online-Ausg.] ed.). Hoboken, NJ: John Wiley & Sons. p. 464. ISBN 0-471-79147-4.
- Floyd, Michael D.; Hillman, Garth D. (8 October 2018) [1st pub. 2000]. "Pulse-Code Modulation Codec-Filters". *The Communications Handbook* (2nd ed.). CRC Press. pp. 26–1, 26–2, 26–3.
- Allstot, David J. (2016). "Switched Capacitor Filters" (PDF). In Maloberti, Franco; Davies, Anthony C. (eds.). *A Short History of Circuits and Systems: From Green, Mobile, Pervasive Networking to Big Data Computing*. IEEE Circuits and Systems Society. pp. 105–110. ISBN 9788793609860.
- Gray, Robert M. (2010). "A History of Realtime Digital Speech on Packet Networks: Part II of Linear Predictive Coding and the Internet Protocol" (PDF). *Found. Trends Signal Process.* 3 (4): 203–303
- Schroeder, Manfred R. (2014). "Bell Laboratories". *Acoustics, Information, and Communication: Memorial Volume in Honor of Manfred R. Schroeder*. Springer. p. 388.
- Janssens, Jelle; Stijn Vandaele; Tom Vander Beken (2009). "The Music Industry on (the) Line? Surviving Music Piracy in a Digital Era". *European Journal of Crime, Criminal Law and Criminal Justice*. 77 (96): 77–96
- Pérez-Bellido A, Soto-Faraco S, López-Moliner J. Sound-driven enhancement of vision: disentangling detection-level from decision-level contributions. *Journal of Neurophysiology*. 2013;109(4):1065–77. pmid:23221404
- Yang W, Li Q, Ochi T, Yang J, Gao Y, Tang X, et al. Effects of auditory stimuli in the horizontal plane on audiovisual integration: an event-related potential study. *PloS one*. 2013;8(6):e66402. pmid:23799097
- Lippert M, Logothetis NK, Kayser C. Improvement of visual contrast detection by a simultaneous sound. *Brain Research*. 2007;1173(0):102–9. pmid:17765208
- Wu J, Yang W, Gao Y, Kimura T. Age-related multisensory integration elicited by peripherally presented audiovisual stimuli. *NeuroReport*. 2012;23(10):616–20. pmid:22643234
- Copeland, Peter (2008). *Manual of Analogue Audio Restoration Techniques* (PDF). London: The British Library. pp. 89–90. Archived (PDF) from the original on December 22, 2015. Retrieved December 16, 2015.
- Koetsier, Teun (2001). "On the prehistory of programmable machines: musical automata, looms, calculators". *Mechanism and Machine Theory*. Elsevier. 36 (5): 589–603. doi:10.1016/S0094-114X(01)00005-2.

- Mitchell, Thomas (2006). Rosslyn Chapel: The Music of the Cubes. Diversions Books. ISBN 0-9554629-0-8.
- Anderson, T.W., 2003. An Introduction to Multivariate Statistical Analysis. Wiley, New York.
- Pitsikalis, V., Kokkinos, I., Maragos, P., 2003. Nonlinear analysis of speech signals: generalized dimensions and Lyapunov exponents. In: Proceedings of the European Conference on Speech Communication and Technology, Eurospeech-03, Geneva, Switzerland, pp. 817–820.
- Potamianos, G., Neti, C., Luettin, J., Matthews, I., 2004. Audio-visual automatic speech recognition: an overview. In: Bailly, G., Vatikiotis-Bateson, E., Perrier, P. (Eds.), Issues in Visual and Audio-Visual Speech Processing. MIT Press (Chapter 10)
- P. B. Denes and E. N. Pinson, The speech chain. W. H. Freeman Company, 2nd Edition, 1993.
- J. L. Flanagan, Speech Analysis, Synthesis and Perception. Springer-Verlag, 1972.
- Raczynski, “Speech processing algorithm for isolated words recognition,” 2018 International Interdisciplinary PhD Workshop (IIPhDW), Swinoujście, 2018, pp. 27-31.
- Z. Muhammad, M. Nasrun, C. Setianingsih and M. A. Murti, “Speech recognition for English to Indonesian translator using hidden Markov model,” 2018 International Conference on Signals and Systems (ICSigSys), Bali, 2018, pp. 255-260.
- Upadhyaya, S. K. Mittal, Y. V. Varshney, O. Farooq and M. R. Abidi, “Speaker adaptive model for Hindi speech using Kaldi speech recognition toolkit,” 2017 International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT), Aligarh, 2017, pp. 222-226.
- Kiss and K. Vicsi, “Investigation of cross-lingual depression prediction possibilities based on speech processing,” 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Debrecen, 2017, pp. 000097-000102.
- K. Gowda, V. Nimbalkar, R. Lavanya, S. Lalitha and S. Tripathi, “Affective computing using speech processing for call centre applications,” 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udipi, 2017, pp. 766-771.
- R. L. Daalache, D. Addou and M. Boudraa, “An efficient distributed speech processing in noisy mobile communications,” 2017 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), Fez, 2017, pp. 1-4.
- Wolpaw, J., Birbaumer, N., McFarland, D., Pfurtscheller, G., & Vaughan, T.

(2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6), 767-791.

Pfurtscheller, G., & Neuper, C. (2001). Motor imagery and direct brain-computer Communication. *Proceedings of the IEEE*, 89(7), 1123-1134.

Millán, J. (2008). Brain-controlled robots. *IEEE Intelligent Systems*

Leeb, R., Scherer, R., Friedman, D., Lee, F., Keinrath, C., Bischof, H., et al. (2007). Towards brain-computer interfacing. In (G. Dornhege, R. Millan Jdel, T. Hinterberger, D. J. McFarland & K. R. Müller ed., chap. Combining BCI and Virtual Reality: Scouting Virtual Worlds). MIT Press.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4, R1-R13.

Lotte, F., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). Studying the use of fuzzy inference systems for motor imagery classification. *IEEE Transactions on Neural System and Rehabilitation Engineering*, 15(2), 322-324.

Lotte, F., Renard, Y., & Lécuyer, A. (2008). Self-paced brain-computer interaction with virtual worlds: a qualitative and quantitative study 'out-of-the-lab'. In 4th international brain-computer interface workshop and training course (p. 373-378).

Neuper, C., Scherer, R., Wriessnegger, S., & Pfurtscheller, G. (2009). Motor imagery and action observation: Modulation of sensorimotor brain rhythms during mental control of a braincomputer interface. *Clinical Neurophysiology*, 120(2), 239-247.

- Bashashati, A., Fatourehchi, M., Ward, R. K., & Birch, G. E. (2007). A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural engineering*, 4(2), R35-57.
- Witten, Ian H.; Frank, Eibe; Hall, Mark A.; Pal, Christopher J. (2011). "Data Mining: Practical machine learning tools and techniques, 3rd Edition". Morgan Kaufmann, San Francisco (CA). Retrieved 2011-01-19.
- Holmes, Geoffrey; Donkin, Andrew; Witten, Ian H. (1994). "Weka: A machine learning workbench" (PDF). *Proceedings of the Second Australia and New Zealand Conference on Intelligent Information Systems*, Brisbane, Australia. Retrieved 2007-06-25.
- Garner, Stephen R.; Cunningham, Sally Jo; Holmes, Geoffrey; Nevill-Manning, Craig G.; Witten, Ian H. (1995). "Applying a machine learning workbench: Experience with agricultural databases" (PDF). *Proceedings of the Machine Learning in Practice Workshop, Machine Learning Conference*, Tahoe City (CA), USA. pp. 14–21. Retrieved 2007-06-25.
- Reutemann, Peter; Pfahringer, Bernhard; Frank, Eibe (2004). "Proper: A Toolbox for Learning from Relational Data with Propositional and Multi-Instance Learners". *17th Australian Joint Conference on Artificial Intelligence (AI2004)*. Springer-Verlag. CiteSeerX 10.1.1.459.8443
- Witten, Ian H.; Frank, Eibe; Trigg, Len; Hall, Mark A.; Holmes, Geoffrey; Cunningham, Sally Jo (1999). "Weka: Practical Machine Learning Tools and Techniques with Java Implementations" (PDF). *Proceedings of the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems*. pp. 192–196. Retrieved 2007-06-26.
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1023/A:1022627411411
- Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comp.* 14, 2531–2560. doi: 10.1162/089976602760407955
- Gallicchio, C., Martin-Guerrero, J. D., Micheli, A., and Soria-Olivas, E. (2017). "Randomized machine learning approaches: Recent developments and challenges," in *Proceedings of the 25th European Symposium on Artificial Neural Networks (ESANN)* (Bruges), 77–86.
- Orr, M. J. L. (1996). *Introduction to Radial Basis Function Networks*. Technical Report, Center for Cognitive Science, University of Edinburgh, Edinburgh, UK.
- Lipton, Z. C. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. *CoRR*. (2015) arXiv:1506.00019.
- Russell, S. J., and Peter, N. (2016). *Artificial Intelligence: A Modern Approach*. Malaysia: Pearson Education Limited.

- Watkins, C. J. C. H., and Dayan, P. (1992). Q-learning. *Mach. Learn.* 8, 279–292. doi: 10.1007/BF00992698
- Schwarzer, M., Rogan, B., Ruan, Y., Song, Z., Lee, D., Percus, A. G., et al. (2019). Learning to fail: predicting fracture evolution in brittle materials using recurrent graph convolutional neural networks. *Comput Mater Sci.* 162, 322–332. doi: 10.1016/j.commatsci.2019.02.046
- Chang, Y.-W., and Lin, C.-J. (2008). “Feature ranking using linear SVM,” in *Causation and Prediction Challenge*, eds I. Guyon, C. Aliferis, G. Cooper, A. Elisseeff, J.-P. Pellet, P. Spirtes, and A. Statnikov, *Proceedings of Machine Learning Research (PMLR)* (Taipei), 53–64.
- Cecen, A., Dai, H., Yabansu, Y. C., Kalidindi, S. R., and Le, S. (2018). Material structure-property linkages using three-dimensional convolutional neural networks. *Acta Mater.* 146, 76–84. doi: 10.1016/j.actamat.2017.11.053
- Lubbers, N., Lookman, T., and Barros, K. (2017). Inferring low-dimensional microstructure representations using convolutional neural networks. *Phys. Rev. E* 96:52111. doi: 10.1103/PhysRevE.96.052111
- Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117. doi: 10.1016/j.neunet.2014.09.003
- LeCun, Y., Haffner, P., Bottou, L., and Bengio, Y. (1999). “Object Recognition with Gradient-Based Learning,” in *Shape, Contour and Grouping in Computer Vision* (Berlin, Heidelberg: Springer Berlin Heidelberg), 319–345.
- Fukushima, K. (1980). Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybernetics* 36, 193–202. doi: 10.1007/BF00344251
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735
- Haykin, S. S. (1998). *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice Hall.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* 65, 386–408. doi: 10.1037/h0042519
- Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Netw.* 4, 251–257. doi: 10.1016/0893-6080(91)90009-T
- Fisher, R. A. (1936). "The Use of Multiple Measurements in Taxonomic Problems" (PDF). *Annals of Eugenics.* 7 (2): 179–188. doi:10.1111/j.1469-1809.1936.tb02137.x. hdl:2440/15227.
- McLachlan, G. J. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley Interscience. ISBN 978-0-471-69115-0. MR 1190469.
- Analyzing Quantitative Data: An Introduction for Social Researchers, Debra Wetcher-Hendricks, p.288

- Martinez, A. M.; Kak, A. C. (2001). "PCA versus LDA" (PDF). *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 23 (=2): 228–233. doi:10.1109/34.908974.
- Abdi, H. (2007) "Discriminant correspondence analysis." In: N.J. Salkind (Ed.): *Encyclopedia of Measurement and Statistic*. Thousand Oaks (CA): Sage. pp. 270–275.
- Perriere, G.; Thioulouse, J. (2003). "Use of Correspondence Discriminant Analysis to predict the subcellular location of bacterial proteins". *Computer Methods and Programs in Biomedicine*. 70 (2): 99–105. doi:10.1016/s0169-2607(02)00011-1. PMID 12507786.
- BÖKEOĞLU ÇOKLUK, Ö, & BÜYÜKÖZTÜRK, Ş. (2008). Discriminant function analysis: Concept and application. *Eğitim araştırmaları dergisi*, (33), 73-92.
- Cortes, Corinna; Vapnik, Vladimir N. (1995). "Support-vector networks" (PDF). *Machine Learning*. 20 (3): 273–297. CiteSeerX 10.1.1.15.9362. doi:10.1007/BF00994018. S2CID 206787478.
- Ben-Hur, Asa; Horn, David; Siegelmann, Hava; Vapnik, Vladimir N. ""Support vector clustering" (2001);". *Journal of Machine Learning Research*. 2: 125–137
- S.B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", *Informatica* 31(2007) 249-268, 2007
- Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.
- Ian H. Witten; Eibe Frank; Mark A. Hall (2011). "Data Mining: Practical machine learning tools and techniques, 3rd Edition". Morgan Kaufmann, San Francisco. p. 191.
- Yang, J. J.; et al. (2008). "Memristive switching mechanism for metal/oxide/metal nanodevices". *Nat. Nanotechnol.* 3 (7): 429–433. doi:10.1038/nnano.2008.160. PMID 18654568.
- Cuntz, Hermann (2010). "PLoS Computational Biology Issue Image | Vol. 6(8) August 2010". *PLOS Computational Biology*. 6 (8): ev06.i08. doi:10.1371/image.pcbi.v06.i08
- Hopfield, J. J. (1982). "Neural networks and physical systems with emergent collective computational abilities". *Proc. Natl. Acad. Sci. U.S.A.* 79 (8): 2554–2558. Bibcode:1982PNAS...79.2554H. doi:10.1073/pnas.79.8.2554. PMC 346238. PMID 6953413