



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Χρήση βαθιών νευρωνικών δικτύων και
πολλαπλοτήτων για την εκπαίδευση ακουστικού
μοντέλου για αυτόματη αναγνώριση φωνής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΙΩΑΝΝΗ Μ. ΧΑΛΚΙΑΔΑΚΗ

Επιβλέπων: Αλέξανδρος Ποταμιάνος
Καθηγητής Ε.Μ.Π.

Αθήνα, ΗΜΕΡΟΜΗΝΙΑ ΕΞΕΤΑΣΗΣ



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής

Χρήση βαθιών νευρωνικών δικτύων και
πολλαπλοτήτων για την εκπαίδευση ακουστικού
μοντέλου για αυτόματη αναγνώριση φωνής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΙΩΑΝΝΗ Μ. ΧΑΛΚΙΑΔΑΚΗ

Επιβλέπων: Αλέξανδρος Ποταμιάνος
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την ΗΜΕΡΟΜΗΝΙΑ ΕΞΕΤΑΣΗΣ.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Αλέξανδρος Ποταμιάνος
Καθηγητής Ε.Μ.Π.

.....
Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

.....
Shrikanth Narayanan
Professor U.S. California

Αθήνα, ΗΜΕΡΟΜΗΝΙΑ ΕΞΕΤΑΣΗΣ



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής

Copyright ©–All rights reserved. Χαλκιαδάκης Μ. Ιωάννης, 2016.

Με την επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Το περιεχόμενο και τα συμπεράσματα αυτής της εργασίας εκφράζουν το συγγραφέα και όχι το Εθνικό Μετσόβιο Πολυτεχνείο ή την εξεταστική επιτροπή.

Περίληψη

Ο στόχος της παρούσας μελέτης ήταν να μελετήσει αρχιτεκτονικές βαθιών νευρωνικών δικτύων οι οποίες έχουν λάβει τεράστια προσοχή κατά τη διάρκεια των τελευταίων ετών, λόγω της επιτυχίας τους σε εφαρμογές που ενδιαφέρουν την επιστημονική κοινότητα μηχανικής μάθησης.

Το πεδίο εφαρμογής που επιλέξαμε ήταν η αυτόματη αναγνώριση φωνής, δεδομένου ότι οι περισσότερες ανακαλύψεις στα βαθιά νευρωνικά δίκτυα έχουν παρουσιάστηκαν για πρώτη φορά σε εφαρμογές αναγνώρισης φωνής. Επιπλέον, υιοθετήσαμε μια προσέγγιση με χρήση πολλαπλότητας για την βελτίωση του κριτηρίου εκπαίδευσης του δικτύου. Η ιδέα (Tomar και Rose, 2014) είναι ότι αν καταφέρουμε να διατηρήσουμε, μέσω του δικτύου, τις σχέσεις των δεδομένων εισόδου που επιβάλλονται από τη δομή της πολλαπλότητας, θα μάθουμε μια πιο ακριβή και εύρωστη κατανομή των κλάσεων φωνημάτων που βρίσκονται στα δεδομένα εισόδου. Ο αλγόριθμος που θα διατηρήσει τις σχέσεις των δεδομένων εισόδου που επιβάλλονται από τη δομή της πολλαπλότητας, χρησιμοποιεί τις κλάσεις φωνημάτων και τις αποστάσεις μεταξύ των χαρακτηριστικών της φωνής για να μάθει την υποκείμενη πολλαπλότητα.

Το πρώτο κεφάλαιο παρουσιάζει το απαραίτητο μαθηματικό υπόβαθρο για να εισαχθεί η έννοια της πολλαπλότητας, προκειμένου να βοηθηθεί ο αναγνώστης με την κατανόηση της πολλαπλότητας και το λόγο για τον οποίο είναι χρήσιμη για την εκπαίδευση του συστήματος.

Στη συνέχεια, προχωράμε παρουσιάζοντας μια γενική, σύντομη επισκόπηση της αυτόματης αναγνώρισης φωνής, καθώς και τις τρέχουσες προσεγγίσεις στον τομέα.

Το τρίτο κεφάλαιο είναι μια εισαγωγή στην μάθηση πολλαπλότητας, όπου ορίζουμε την περιοχή και δίνουμε μια γενική εικόνα σχετικών δημοφιλών αλγορίθμων. Στην τελευταία ενότητα αυτού του κεφαλαίου, παρουσιάζουμε τη δουλειά των Τομαρ και Ροσε, 2014, στην οποία βασίστηκε το τρέχον έργο.

Στο τέταρτο κεφάλαιο παρουσιάζουμε τα βαθιά νευρωνικά δίκτυα ξεκινώντας την ιστορία και τα κίνητρα πίσω από τις βαθιές αρχιτεκτονικές στην αναγνώριση φωνής και προχωρώντας με την περιγραφή της βαθιάς αρχιτεκτονικής περσεπτρον, την οποία χρησιμοποιήσαμε στην παρούσα εργασία. Δίνονται πρακτικές συμβουλές όπως διαπιστώθηκαν κατά τη διάρκεια της μελέτης και του πειραματικού μέρους της εργασίας. Το κεφάλαιο ολοκληρώνεται με την σύντομη παρουσίαση των ρεσυρρεντ αρχιτεκτονικών, ένα είδος δικτύου που έχει λάβει πολλή προσοχή από την επιστημονική κοινότητα μηχανικής μάθησης.

Το τελευταίο κεφάλαιο περιγράφει λεπτομερώς τη δουλειά που πραγματοποιήσαμε, τον τρόπο που ενσωματώσαμε την πολλαπλότητα στο βαθύ νευρωνικό δίκτυο, καθώς και τις

προκλήσεις που αντιμετωπίσαμε κατά τη διάρκεια της εργασίας. Τέλος, παρουσιάζονται τα πειραματικά αποτελέσματα και επακόλουθες παρατηρήσεις.

Λέξεις Κλειδιά

βαθιά νευρωνικά δίκτυα, μηχανική μάθηση, μάθηση πολλαπλοτήτων, κοντινότεροι γείτονες, αυτόματη αναγνώριση φωνής, αυτόματη αναγνώριση φωνής συνεχούς λόγου και μεγάλου λεξιλογίου, ακουστικό μοντέλο, υβριδικό ακουστικό μοντέλο

