

Έξυπνο, εξειδικευμένο περιβαλλοντικό παρατηρητήριο στην Μεσσηνία

Παραδοτέο 8.1

Ανάλυση δεδομένων, βελτιστοποίηση των αλγορίθμων



Δρ. Χρήστος Μάρκου, Διευθυντής Ερευνών, ΙΠΣΦ - ΕΚΕΦΕ "Δημόκριτος"

Δημήτριος Σταυρόπουλος, Υποψήφιος Διδάκτορας, ΙΠΣΦ - ΕΚΕΦΕ "Δημόκριτος"

Αθήνα, Σεπτέμβριος 2023

Περιεχόμενα

Περιεχόμενα	2
1 Βελτιστοποίηση αλγορίθμων για την κατηγοριοποίηση ηχητικών παλμών	3
1.1 Τεχνικές πληροφορίες	3
1.2 Αρχιτεκτονική Νευρωνικού Δικτύου	4
1.3 Αξιολόγηση αποτελεσμάτων	5



ΕΥΡΩΠΑΪΚΗ ΕΝΩΣΗ
Ευρωπαϊκό Ταμείο
Περιφερειακής Ανάπτυξης

ΠΕΡΙΦΕΡΕΙΑΚΟ ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
«ΠΕΛΟΠΟΝΝΗΣΟΣ 2014-2020»



Κεφάλαιο 1

Βελτιστοποίηση αλγορίθμων για την κατηγοριοποίηση ηχητικών παλμών

Οι αλγορίθμοι που περιγράφονται στο παρόν κεφάλαιο αναπτύχθηκαν με σκοπό την βελτιστοποίηση της διαδικασίας ταξινόμησης ηχητικών σημάτων, όπως περιγράφηκε στο ΠΕ5 του έργου. Ο νετρινικός διπολικός παλμός θεωρείται ομοίως σήμα ενδιαφέροντος, με βάση τα ίδια κριτήρια που αναφέρθηκαν στην παράγραφο 2.2.1 του ΠΕ5. Για την ανάπτυξη αλγορίθμων με σκοπό την κατηγοριοποίηση των ηχητικών παλμών, γίνεται χρήση των ίδιων κλάσεων σήματος και υποβάθρου όπως και προεργασία των chunks όπως περιγράφηκαν στις παραγράφους 3.1 και 3.2 του ΠΕ5, με σκοπό η σύγκριση που προκύπτει ως προς την απόδοση τους να είναι δόκιμη.

1.1 Τεχνικές πληροφορίες

Για την ταξινόμηση των chunks χρησιμοποιείται ένα Νευρωνικό Δίκτυο με περισσότερα από δύο hidden layers (Deep Neural Network). Συγκεκριμένα το Deep Neural Network που χρησιμοποιείται ανήκει στην κατηγορία των Convolutional Neural Networks, τύπος που χρησιμοποιείται κατά κόρον σε εφαρμογές αναγνώρισης εικόνας και ήχου. Για την εκτέλεση τέτοιου είδους αλγορίθμων σε ένα εύλογο χρονικό διάστημα, απαιτείται ικανή υπολογιστική ισχύς. Στο πλαίσιο του ISEO δημιουργήθηκε ένας cluster στον server του Ινστιτούτου Πυρηνικής και Σωματιδιακής Φυσικής του ΕΚΕΦΕ "Δημόκριτος". Η state of the art λύση που ακολουθήθηκε ήταν να εφοδιαστεί ο συγκεκριμένος cluster με GPUs. Η χρήση τέτοιων μονάδων παρέχει αρκετά ταχύτερη επεξεργασία σε περιστάσεις όπως η προώθηση πληροφορίας μεταξύ των στρωμάτων ενός Deep Neural Network, σε περιπτώσεις δηλαδή που η πληροφορία πρέπει να επεξεργαστεί ή να προωθηθεί χωρίς ιδιαίτερη πολυπλοκότητα.

1.2 Αρχιτεκτονική Νευρωνικού Δικτύου

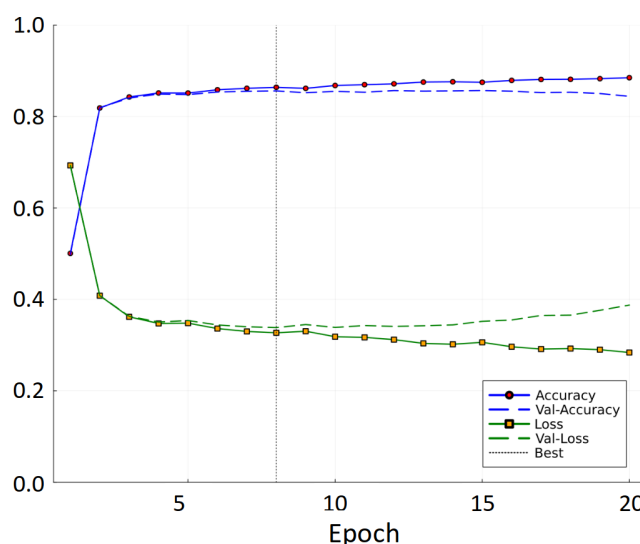
Το Convolutional Neural Networks που χρησιμοποιείται για την ηχητική κατηγοριοποίηση αποτελείται από διάφορα επίπεδα (layers), όπου το καθένα εξυπηρετεί μία συγκεκριμένη λειτουργία. Οι τύποι των επιπέδων έχουν ως εξής:

- **Εισαγωγικό επίπεδο (input layer).** Το εισαγωγικό επίπεδο αποτελείται από 720 νευρώνες, όσες είναι και οι αριθμητικές τιμές που είναι αποθηκευμένες σε κάθε ακουστικό αρχείο.
- **Μονοδιάστατα επίπεδα συσχέτισης (1D convolutional layer).** Αυτά τα επίπεδα σαρώνουν γειτονιές νευρώνων στα ακριβώς προηγούμενα επίπεδα, εφαρμόζουν φίλτρα συσχέτισης και εξάγουν χαρακτηριστικά (features). Το μέγεθος των γειτονιών, το μέγεθος των φίλτρων και το πλήθος των χαρακτηριστικών είναι υπερπαράμετροι που πρέπει να καθοριστούν συγκεκριμένα σε κάθε μονοδιάστατο επίπεδο συσχέτισης με στρατηγική, λαμβάνοντας υπόψιν το πλήθος των δεδομένων, τη φυσική τους σημασία αλλά και εξετάζοντας την τελική απόδοση ολόκληρου του μοντέλου.
- **Μονοδιάστατα επίπεδα μέγιστης συγκέντρωσης (1D max-pooling layers).** Αυτά χρησιμοποιούνται αμέσως μετά από μονοδιάστατα επίπεδα συσχέτισης, όπου θα εφαρμόσουν φίλτρα πάνω στα χαρακτηριστικά για να μειώσουν το δείγμα των δεδομένων. Η χρησιμότητα σε αυτό είναι ότι το συνολικό μοντέλο δεν είναι εξαιρετικά ευαίσθητο σε μια συγκεκριμένη κατάσταση των χαρακτηριστικών που εντοπίζει, καθώς και αποφεύγουμε την υπερ-εκπαίδευση του μοντέλου (over-training). Έτσι το μοντέλο βοηθάται να εντοπίζει τα σημαντικότερα χαρακτηριστικά, μειώνοντας ταυτόχρονα τις υπολογιστικές απαιτήσεις στα επόμενα επίπεδα.
- **Πυκνά επίπεδα (dense layers).** Τοποθετούνται στο τέλος ενός μοντέλου, όταν έχουν εξαχθεί αρκετά χαρακτηριστικά. Όλοι οι νευρώνες ενός πυκνού επιπέδου συνδέονται με όλους τους νευρώνες του αμέσως προηγούμενου επιπέδου, ώστε να επιτύχουν την τελική πρόβλεψη και κατηγοροποίηση των δεδομένων.
- **Επίπεδα απόπτωσης (dropout layers).** Ο στόχος αυτών των επιπέδων είναι να προστατέψουν το μοντέλο από την υπερ-εκπαίδευση. Το επιτυγχάνουν με την εφαρμογή τους στο αποτέλεσμα των πυκνών επιπέδων, όπου απενεργοποιείται ένα συγκεκριμένο ποσοστό τυχαία επιλεγμένων νευρώνων. Οι απενεργοποιημένοι νευρώνες δεν συμμετέχουν στην εμπρόσθια και οπίσθια μεταφορά πληροφορίας κατά την εκπαίδευση του μοντέλου και έτσι το μοντέλο δεν βασίζεται μόνο σε συγκεκριμένους νευρώνες για την τελική πρόβλεψη.
- **Εξαγωγικό επίπεδο (output layer).** Περιέχει τόσους νευρώνες όσες είναι και οι κλάσεις κατηγοροποίησης (2 κλάσεις στο συγκεκριμένο μοντέλο).

Η αρχιτεκτονική που χρησιμοποιήθηκε αποτελείται από ένα εισαγωγικό επίπεδο, τρία μονοδιάστατα επίπεδα συσχέτισης και τρία μονοδιάστατα επίπεδα μέγιστης συγκέντρωσης τοποθετημένα εναλλάξ, ένα πυκνό επίπεδο 512 νευρώνων, ένα επίπεδο απόπτωσης 50% (απενεργοποίηση των μισών νευρώνων), και ένα εξαγωγικό επίπεδο.

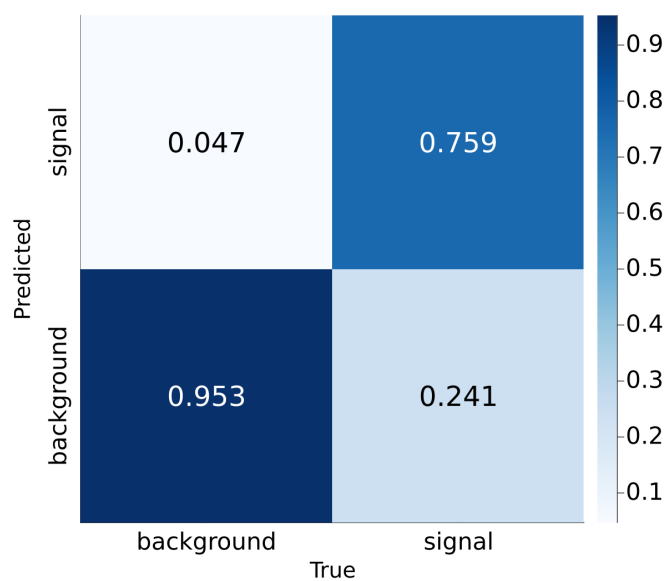
1.3 Αξιολόγηση αποτελεσμάτων

Η εκπαίδευση του μοντέλου πραγματοποιήθηκε με τα chunks που κατασκευάστηκαν για τις δύο κλάσεις, εκτός από ένα ποσοστό το οποίο παρέμεινε άγνωστο στο μοντέλο, και χρησιμοποιήθηκε για τον έλεγχο της επίδοσής του. Το *accuracy*, ορισμένο ως το ποσοστό των chunks τα οποία ταξινομούνται σωστά, ως προς το πλήθος του συνόλου των chunks, παρουσιάζεται ως συνάρτηση των επαναλήψεων εκπαίδευσης (σχήμα 1.1). Για το *accuracy* παρατηρούνται ποσοστά της τάξεως του $\sim 85\%$.



Σχήμα 1.1: Accuracy συνάρτησε των επαναλήψεων εκπαίδευσης.

Η αποτελεσματικότητα του αλγορίθμου αξιολογείται επιπροσθέτως από το confusion matrix του σχήματος 1.2. Παρατηρείται ότι το 75.9% από τα chunks του σήματος που εξαιρέθηκαν για το δείγμα ελέγχου αξιολογούνται σωστά. Το ίδιο ισχύει για το 95.3% των chunks υποβάθρου. Το 24.1% των chunks σήματος αξιολογούνται ως υπόβαθρο, ενώ μόνο το περίπου 4.7% των chunks υποβάθρου αξιολογούνται ως σήμα. Συνεπώς η απόρριψη του υποβάθρου κυμαίνεται στο επίπεδο του $\sim 96\%$, ενώ παράλληλα το ποσοστό των chunks που αξιολογείται σωστά αυξήθηκε κατά $\sim 7\%$.



Σχήμα 1.2: Confusion matrix για την ταξινόμηση των chunks σήματος και υποβάθρου.