ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μάθημα Προπτυχιακών Σπουδών:

Επεξεργασία Σημάτων Φωνής και Ήχου

Ακαδημαϊκό Έτος: 2024 – 2025

Εξάμηνο: 10ο

Απαλλακτική Εργασία

<u>Στοιχεία Φοιτητή</u>

Ονοματεπώνυμο: Ζηνοβία Γκούμα

ΑΜ: Π20048

Περιεχόμενα

Εισαγωγή 3
Ορισμός συναρτήσεων 3
Κυρίως πρόγραμμα 5
Παράδειγμα εκτέλεσης 5
Συμπεράσματα 8
Πηγές 8

Εισαγωγή

Για την επιτυχής εκτέλεση του προγράμματος, πρέπει να εγκαταστήσετε τις βιβλιοθήκες που υπάρχουν μέσα στο αρχείο requirements.txt**.**

Η εκπόνηση της εργασίας έγινε με την γλώσσα προγραμματισμού Python (v 3.12.0).

Στόχος της παρούσας εργασίας είναι η δημιουργία ενός προγράμματος, το οποίο χωρίζει μια πρόταση σε τμήματα σήματος ομιλίας και σήματος υποβάθρου. Ως αποτέλεσμα, δίνοντας στο πρόγραμμα ένα αρχείο ήχου, αυτό θα επιστρέφει τα χρονικά όρια των τμημάτων σήματος ομιλίας και των τμημάτων σήματος υποβάθρου σε μορφή ενός αρχείου

Για την εκπαίδευση των ταξινομητών, χρησιμοποιήθηκε ένας κατάλογος 'train' που περιέχει αρχεία ομιλίας και αρχεία υποβάθρου. Επειδή το πλήθος των αρχείων υποβάθρου ήταν πολύ μεγάλο, πήρα ένα υποσύνολο αυτού αγνοώντας τα αρχεία που ξεπερνάνε τα 2700 Για την τελική δοκιμή, χρησιμοποιήθηκε ο κατάλογος 'test' ο οποίος περιέχει το αρχείο ήχου "S01_U04.CH4.wav" και τα αρχεία μεταγραφής "S01.json" και "S21.json".

Τα παραπάνω αρχεία μας ήταν διαθέσιμα από την εκφώνηση, αλλά υπάρχουν και στα δημόσια σύνολα δεδομένων Open SLR, MUSAN και CHiME.

Η συνολικός κώδικας αποτελείται από δυο scripts, το *functions.py*, όπου ορίζονται οι κατάλληλες συναρτήσεις και το *main.py* όπου βρίσκεται το κυρίως πρόγραμμα.

Σημείωση: Δεν γίνεται ιδιαίτερη επεξήγηση του κώδικα στο παρόν έγγραφο, διότι υπάρχουν εκτενή σχόλια στον κώδικα

Ορισμός συναρτήσεων

Αρχικά, φορτώνω όλα τα αρχεία ήχου (ομιλίας και υποβάθρου) με την συνάρτηση Έπειτα, με την *frame_signal* τα συνεχή ηχητικά σήματα χωρίζονται σε πλαίσια. Επέλεξα πλαίσια διάρκειας 25ms και hop 10ms, διότι οι συγκεκριμένες τιμές είναι στάνταρ στην ανάλυση ομιλίας και εξασφαλίζουν καλή ισορροπία ανάμεσα στην χρονική και φασματική ανάλυση. Έτσι, επιστρέφεται ένας πίνακας όπου κάθε γραμμή είναι ένα πλαίσιο. Αφού ολοκληρωθεί και η εξαγωγή διανυσμάτων χαρακτηριστικών για κάθε πλαίσιο μέσω της προχωράμε στην δημιουργία του συνόλου δεδομένων για την εκπαίδευση των ταξινομητών.

Στην *create_dataset*, για κάθε πλαίσιο του σήματος ανάλογα τα χαρακτηριστικά του αντιστοιχείται σε μια ετικέτα. Τελικά, επιστρέφονται όλα τα εξαγόμενα χαρακτηριστικά και οι αντίστοιχες ετικέτες ως numpy arrays.

Προχωράμε στην εκπαίδευση των ταξινομητών, η οποία γίνεται στις συναρτήσεις *LSQ_train* και *MLP_train* για τους Least Squares και MLP αντίστοιχα.

Εκπαίδευση ταξινομητή Least Squares

Παίρνει μέρος στην συνάρτηση *LSQ_train* και εκπαιδεύει ένα μοντέλο γραμμικής παλινδρόμησης (Least Squares) με βάση τα δεδομένα X (χαρακτηριστικά) και y (ετικέτες). Χρησιμοποιεί τη LinearRegression() από τη βιβλιοθήκη *sklearn*.

Εκπαίδευση ταξινομητή MLP (Multi-Layer Perceptron)

Πραγματοποιείται στην συνάρτηση *MLP_train* και βασίζεται πάλι στα δεδομένα X και y. Με την βοήθεια της κλάσης MLPClassifier() της βιβλιοθήκης *sklearn*, εκπαιδεύει ένα νευρωνικό δίκτυο που έχει τρία κρυφά επίπεδα 64, 32 και 16 νευρώνων αντίστοιχα και μέχρι 500 επαναλήψεις εκπαίδευσης.

Στην συνέχεια, αφού με την *predict_segments* προβλέψουμε σε ποια κατηγορία ανήκει το κάθε πλαίσιο (ομιλία ή υπόβαθρο), με την *get_segments* θα εντοπίσουμε τα χρονικά διαστήματα όπου εμφανίζεται μια συγκεκριμένη ετικέτα σε μορφή timestamps. Επίσης, οι συναρτήσεις *get_speech_intervals* και *get_noise_intervals* δουλεύουν με τις χειροκίνητες σημειώσεις (annotations) στα ηχητικά αρχεία και επιστρέφουν μια λίστα από ζεύγη (start, που στην πρώτη είναι τα διαστήματα ομιλίας και στην δεύτερη τα διαστήματα υποβάθρου. Η συνάρτηση *generate_ground_truth_labels* φτιάχνει μια λίστα με ετικέτες (0 ή 1) για κάθε frame του ήχου, με βάση τα χρονικά διαστήματα που υπάρχει ομιλία (speech_intervals).

Τέλος, με την *save_csv* τα αποτελέσματα του προγράμματος αποθηκεύονται με το επιθυμητό format σε ένα αρχείο csv.

Κυρίως πρόγραμμα

Στο κυρίως πρόγραμμα, λαμβάνει χώρα ο έλεγχος τον ταξινομητών. Αρχικά εμφανίζεται ένα μενού που λειτουργεί ως διεπαφή με τον χρήστη και του δίνει επιλογές για τον τρέξιμο του προγράμματος. Αφού χρήστης επιλέξει τον ταξινομητή που επιθυμεί, εμφανίζονται στη κονσόλα οι ενέργειες της διαδικασίας που γίνεται και το ποσοστό ακρίβειας του μοντέλου που επιλέχθηκε.

Παράδειγμα εκτέλεσης

```
=== Ομιλία και υπόβαθρο ===
Διάλεξε ενέργεια:
1 - Χρήση Least Squares
2 - Χρήση MLP
3 - Έξοδος από το πρόγραμμα
Εισήγαγε επιλογή:
```

Μόλις τρέξει το πρόγραμμα εμφανίζεται ένα μενού που παρακινεί τον χρήστη να διαλέξει ενέργεια. Οι διαθέσιμες ενέργειες είναι:

1. Χρήση μοντέλου Least Squares
2. Χρήση μοντέλου MLP
3. Έξοδος από το πρόγραμμα

Ποσοστό ακρίβειας Least Squares και αρχείο csv

```
=== Ομιλία και υπόβαθρο ===
Διάλεξε ενέργεια:
1 - Χρήση Least Squares
2 - Χρήση MLP
3 - Έξοδος από το πρόγραμμα
Εισήγαγε επιλογή: 1
Φόρτωση αρχείων...
Δημιουργία συνόλου δεδομένων...
Εκπαίδευση LSQ...

Πρόβλεψη στο αρχείο: S01_U04.CH4.wav
Ακρίβεια μοντέλου: 52.5%

Τα αποτελέσματα αποθηκεύτηκαν στο results.csv

Process finished with exit code 0
```

	A	B	C	D	E
1	Audiofile	start	end	class	
2	S01_U04.C	0	0.23	background	
3	S01_U04.C	0.23	0.52	foreground	
4	S01_U04.C	0.52	0.56	background	
5	S01_U04.C	0.56	1.12	foreground	
6	S01_U04.C	1.12	1.16	background	
7	S01_U04.C	1.16	1.29	foreground	
8	S01_U04.C	1.29	1.35	background	
9	S01_U04.C	1.35	1.37	foreground	
10	S01_U04.C	1.37	1.47	background	
11	S01_U04.C	1.47	1.57	foreground	
12	S01_U04.C	1.57	1.82	background	
13	S01_U04.C	1.82	2.04	foreground	
14	S01_U04.C	2.04	2.39	background	
15	S01_U04.C	2.39	2.44	foreground	
16	S01_U04.C	2.44	2.62	background	
17	S01_U04.C	2.62	3.69	foreground	
18	S01_U04.C	3.69	3.82	background	
19	S01_U04.C	3.82	4.04	foreground	
20	S01_U04.C	4.04	4.09	background	
21	S01_U04.C	4.09	4.12	foreground	
22	S01_U04.C	4.12	4.23	background	
23	S01_U04.C	4.23	4.78	foreground	
24	S01_U04.C	4.78	4.97	background	
25	S01_U04.C	4.97	5.08	foreground	
26	S01_U04.C	5.08	5.11	background	
27	S01_U04.C	5.11	5.19	foreground	
28	S01_U04.C	5.19	5.25	background	
29	S01_U04.C	5.25	5.3	foreground	
30	S01_U04.C	5.3	5.46	background	
31	S01_U04.C	5.46	5.51	foreground	
32	S01_U04.C	5.51	6.07	background	
33	S01_U04.C	6.07	6.13	foreground	
34	S01_U04.C	6.13	6.18	background	
35	S01_U04.C	6.18	6.31	foreground	
36	S01_U04.C	6.31	6.32	background	

Ποσοστό ακρίβειας MLP και αρχείο csv

```
=== Ομιλία και υπόβαθρο ===
Διάλεξε ενέργεια:
1 - Χρήση Least Squares
2 - Χρήση MLP
3 - Έξοδος από το πρόγραμμα
Εισήγαγε επιλογή: 2
Φόρτωση αρχείων...
Δημιουργία συνόλου δεδομένων...
Εκπαίδευση MLP...

Πρόβλεψη στο αρχείο: S01_U04.CH4.wav
Ακρίβεια μοντέλου: 53.4%

Τα αποτελέσματα αποθηκεύτηκαν στο results.csv

Process finished with exit code 0
```

	A	B	C	D	E
1	Audiofile	start	end	class	
2	S01_U04.C	0	0.28	background	
3	S01_U04.C	0.28	0.64	foreground	
4	S01_U04.C	0.64	0.76	background	
5	S01_U04.C	0.76	0.86	foreground	
6	S01_U04.C	0.86	1.51	background	
7	S01_U04.C	1.51	1.52	foreground	
8	S01_U04.C	1.52	1.54	background	
9	S01_U04.C	1.54	1.55	foreground	
10	S01_U04.C	1.55	1.85	background	
11	S01_U04.C	1.85	2.16	foreground	
12	S01_U04.C	2.16	2.7	background	
13	S01_U04.C	2.7	2.89	foreground	
14	S01_U04.C	2.89	2.93	background	
15	S01_U04.C	2.93	3.26	foreground	
16	S01_U04.C	3.26	3.3	background	
17	S01_U04.C	3.3	3.38	foreground	
18	S01_U04.C	3.38	3.4	background	
19	S01_U04.C	3.4	3.41	foreground	
20	S01_U04.C	3.41	3.42	background	
21	S01_U04.C	3.42	3.44	foreground	
22	S01_U04.C	3.44	3.61	background	
23	S01_U04.C	3.61	3.64	foreground	
24	S01_U04.C	3.64	3.65	background	
25	S01_U04.C	3.65	3.66	foreground	
26	S01_U04.C	3.66	3.85	background	
27	S01_U04.C	3.85	4.16	foreground	
28	S01_U04.C	4.16	4.24	background	
29	S01_U04.C	4.24	4.56	foreground	
30	S01_U04.C	4.56	4.63	background	
31	S01_U04.C	4.63	4.7	foreground	
32	S01_U04.C	4.7	4.78	background	
33	S01_U04.C	4.78	5.05	foreground	

Έξοδος προγράμματος

```
=== Ομιλία και υπόβαθρο ===
Διάλεξε ενέργεια:
1 - Χρήση Least Squares
2 - Χρήση MLP
3 - Έξοδος από το πρόγραμμα
Εισήγαγε επιλογή: 3
Έξοδος...

Process finished with exit code 0
```

Συμπεράσματα

Βάσει των ποσοστών ακριβείας των δύο ταξινομητών το οποίο μόλις ξεπερνάει το 50% (δηλαδή την τυχαία επιλογή), συμπεραίνω ότι δεν καταφέρνουν να διακρίνουν αποτελεσματικά μεταξύ ομιλίας και υπόβαθρου με βάση τα δεδομένα και τα χαρακτηριστικά που του έδωσα. Αυτό ίσως οφείλεται στο μέγεθος του συνόλου δεδομένων που χρησιμοποίησα για την εκπαίδευσή τους ή την προσέγγιση που επέλεξα γενικότερα.

Πηγές

- NumPy library
  Χρησιμοποιήθηκε για αριθμητικές πράξεις και διαχείριση πινάκων

- Librosa library
  Χρησιμοποιήθηκε για την ανάγνωση των αρχείων ήχου και την ανάλυσή τους σε πλαίσια, καθώς και για την εξαγωγή διανυσμάτων χαρακτηριστικών

- Scipy library
  Χρησιμοποιήθηκε για την εφαρμογή φίλτρου μέσης τιμής (medfilt) στη σειρά προβλέψεων

- Scikit-learn library
  Χρησιμοποιήθηκε για την εκπαίδευση και αξιολόγηση των ταξινομητών

- OpenSLR
  Από εκεί πήραμε τις συλλογές MUSAN και CHiME για το πρόγραμμα