

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ
Χειμερινό Εξάμηνο 2017-18

Προπαρασκευή 3ης Εργαστηριακής Άσκησης:
Αναγνώριση Συναισθήματος στη Μουσική

ΠΕΡΙΓΡΑΦΗ

Σκοπός είναι η υλοποίηση ενός συστήματος αυτόματης ταξινόμησης μουσικών αποσπασμάτων με βάση το «επιδιώκόμενο» συναίσθημα, δηλαδή το συναίσθημα που το μουσικό κομμάτι προσπαθεί να δημιουργήσει στον ακροατή (*Music Emotion Classification*). Στο πρώτο μέρος (προπαρασκευή) επικεντρωνόμαστε στην εξοικείωση με τα δεδομένα και την εξαγωγή ενός συνόλου χαρακτηριστικών από το μουσικό σήμα, ενώ στο δεύτερο θα διερευνήσουμε την ικανότητά των χαρακτηριστικών αυτών να διαχωρίσουν τις κατηγορίες του συναισθήματος, χρησιμοποιώντας διάφορους αλγορίθμους ταξινόμησης. Στόχος της εργασίας αυτής είναι να εξακριβώσουμε τα χαρακτηριστικά και τις μεθόδους εκείνες που οδηγούν στην καλύτερη ταξινόμηση του συναισθήματος που μεταδίδει η μουσική στον ακροατή.

Συγκεκριμένα, σας δίνεται σύνολο από προσημειωμένα μουσικά αποσπάσματα όπου και θα μελετήσετε την περιγραφή των συναισθημάτων με τη χρήση μιας δισδιάστατης αναπαράστασης του συναισθήματος στους άξονες της Χαράς/Λύπης (*Valence*) και Ενεργοποίησης/Απενεργοποίησης (*Activation*)¹. Τα χαρακτηριστικά θα εξαχθούν με τη βοήθεια του MIRtoolbox. Το τελικό πακέτο χαρακτηριστικών θα περιλαμβάνει ένα συνδυασμό χαρακτηριστικών όπως για παράδειγμα (α) χαρακτηριστικά εμπνευσμένα από τη μουσική και (β) στατιστικά μεγέθη των Mel-Frequency Cepstral Coefficients (MFCCs).

Σημείωση: Για ερωτήσεις επικοινωνήστε με την Επιστημονική Συνεργάτη του μαθήματος Δρ. Νάνσυ Ζλατίντση (nzlat@cs.ntua.gr)

ΕΚΤΕΛΕΣΗ

Χρησιμοποιώντας τα δεδομένα από το link:
http://cvsp.cs.ntua.gr/courses/patrec/labs_material2014/PRcourse_Lab3_data.zip) εκτελέστε τα παρακάτω βήματα χρησιμοποιώντας το Matlab:

¹ J.A. Russell & A. Mehrabian, *Evidence for a three-factor theory of emotions*. Journal of Research in Personality, 11, 273-294, 1977.

Βήμα 1

Προ-επεξεργασία δεδομένων

Η συλλογή δεδομένων η οποία θα χρησιμοποιηθεί αποτελείται από μουσικά αποσπάσματα κομματιών των Beatles (φάκελος: *MusicFileSamples*). Συνολικά θα βρείτε 412 .wav αρχεία, το καθένα με διάρκεια μεταξύ 10-20 δευτερόλεπτα. Η επιλογή τους έγινε έτσι ώστε α) το συναισθηματικό περιεχόμενο να θεωρείται παρόμοιο καθ' όλη τη διάρκεια του αποσπάσματος αλλά και β) να έχουν ομοιογένεια τόσο στην μελωδία όσο και στους στίχους.

Τα ονόματα των αρχείων της βάσης είναι «file1.wav, file2.wav, ...file412.wav». Τα δεδομένα έχουν format: fs = 44.1kHz, 16 bits/sample, stereo.

Εισάγετε τα δεδομένα στο περιβάλλον Matlab, μετατρέψτε τα και αποθηκεύστε τα για τη διαδικασία της εξαγωγής χαρακτηριστικών σε format: fs = 22.05 kHz, 8 bits/sample, mono.

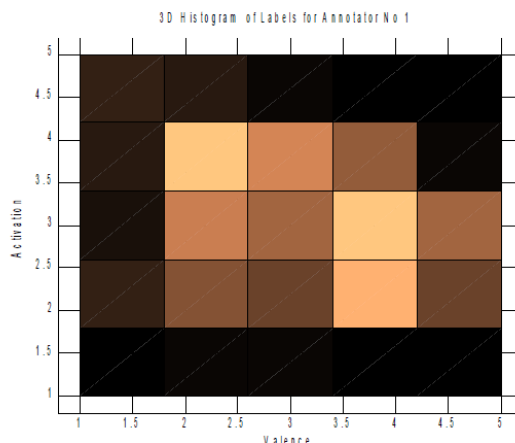
Βήμα 2

Υπολογισμός στατιστικών των δεδομένων επισημείωσης

Για την επισημείωση του συναισθήματος έχει χρησιμοποιηθεί ο διδιάστατος χώρος που περιγράφει τις διαστάσεις **Valence** και **Activation**. Οι επισημειωτές για την περιγραφή του συναισθήματος στην κάθε διάσταση έχουν επιλέξει μία από τις 5 πιθανές ακέραιες τιμές {1,2,3,4,5}. Θεωρούμε πως η τιμή Valence 1 είναι μικρότερη του Valence 2 καθώς και ότι η διαφορά του Activation 2 και Activation 5 είναι μεγαλύτερη από ότι μεταξύ του Activation 4 και Activation 5.

Δίνονται τρία .mat αρχεία (ένα για κάθε επισημειωτή, με ονόματα Labeler1-3, φάκελος: *EmotionLabellingData*) τα οποία περιέχουν τις τιμές του Valence και του Activation {1,2,3,4,5} για κάθε μουσικό απόσπασμα. Εισάγετε τα αρχεία στο περιβάλλον Matlab και υπολογίστε για κάθε επισημειωτή και για τις δύο διαστάσεις α) τη μέση τιμή (mean) και β) την τυπική απόκλιση (standard deviation) και γ) δείξτε τους co-occurrence πίνακες σε 2D ιστογράμματα (Valence Vs. Activation Vs. Number of Occurrences). Ενδεικτικό παράδειγμα οπτικοποίησης για έναν επισημειωτή:

		Valence				
		1	2	3	4	5
A c t i v .	1	0	1	1	0	0
	2	4	10	8	21	8
	3	2	15	12	24	12
	4	3	24	16	11	1
	5	4	3	1	0	0



δ) Σχολιάστε τα αποτελέσματα μετά την επεξεργασία των δεδομένων για κάθε επισημειωτή ξεχωριστά.

Βήμα 3

Η υψηλή συμφωνία μεταξύ των επισημειωτών αποτελεί συνήθως δείκτη πως τα συγκεκριμένα μουσικά αποσπάσματα αποτελούν «καλά παραδείγματα» για την χρήση τους στην εκπαίδευση ενός μοντέλου ταξινόμησης. Σας ζητείται να αξιολογήσετε το ποσοστό συμφωνίας μεταξύ κάθε ζεύγους επισημειωτών.

α) Υπολογίστε την παρατηρούμενη συμφωνία (*Observed Agreement*) καθώς και τη συνολική (κατά μέσο όρο) συμφωνία και στις δύο διαστάσεις για κάθε ζεύγος επισημειωτή χρησιμοποιώντας τον τύπο:

$$1 - E \left\{ \frac{|I_{ij} - I_{ik}|}{4} \right\}$$

όπου I_{ij} είναι η τιμή του επισημειωτή j στο δείγμα i και I_{ik} η τιμή του επισημειωτή k στο δείγμα i . Η διαίρεση με 4 γίνεται γιατί οι τιμές που χρησιμοποιούνται είναι $\{1,2,3,4,5\}$, και άρα η μεγαλύτερη απόλυτη διαφορά μεταξύ I_{ij} και I_{ik} είναι 4.

β) Παρουσιάστε τα ιστογράμματα των διαφορών I_{ij} και I_{ik} κάθε ζεύγους επισημειωτών.

Βήμα 4

Υπολογίστε τον συντελεστή Krippendorff's alpha² (για ordinal δεδομένα) σε κάθε διάσταση. Το alpha μετράει τη συμφωνία μεταξύ των επισημειωτών και δίνεται ως:

$$\alpha = 1 - \frac{\text{Observed Disagreement}}{\text{Expected Disagreement}},$$

όπου η παρατηρούμενη διαφωνία (*Observed Disagreement*) είναι η μέση τιμή όλων των σημείων διαφωνίας των δεδομένων (μετρημένη ως απόλυτη διαφορά) και η αναμενόμενη διαφωνία (*Expected Disagreement*) είναι η μέση τιμή της διαφοράς μεταξύ δύο τιμών i και k σε όλα τα πιθανά ζευγάρια $n(n-1)$ τιμών. Σχολιάστε τα αποτελέσματά σας.

Βήμα 5

Για τον υπολογισμό των τιμών των επισημειώσεων, οι οποίες και θα χρησιμοποιηθούν για την ταξινόμηση των δεδομένων, χρησιμοποιήστε τη μέση τιμή (mean) των τριών επισημειωτών σε κάθε μουσικό απόσπασμα αλλά και για τη κάθε διάσταση ξεχωριστά. Οι τελικές επισημειώσεις θα έχουν τιμές $\{1, 1.3333, 1.6667, 2, 2.3333, 2.6667, 3, 3.3333, 3.6667, 4, 4.3333, 4.6667, 5\}$.

Δείξτε σε 2D ιστόγραμμα τον co-occurrence πίνακα (Valence Vs Activation Vs Number of Occurrences) των τελικών επισημειώσεων.

Βήμα 6

Εξοικείωση με το MIR toolbox και εξαγωγή χαρακτηριστικών.

Για την εξαγωγή των χαρακτηριστικών αρχικά θα χρησιμοποιήσετε βραχείος χρόνου ανάλυση, και στην συνέχεια θα υπολογίσετε στατιστικά που αφορούν την κατανομή των εν λόγω χαρακτηριστικών στο σύνολο των παραθύρων ανάλυσης (π.χ., μέση τιμή, τυπική απόκλιση, μέγιστη και ελάχιστη τιμή κ.ά.).

² Krippendorff, K.: *Computing Krippendorff's Alpha-Reliability*, http://repository.upenn.edu/asc_papers/43/
http://en.wikipedia.org/wiki/Krippendorff%27s_alpha

Για την εξαγωγή των χαρακτηριστικών χρησιμοποιήστε τις προεπιλεγμένες από το MIRtoolbox τιμές, δηλ., διάρκεια πλαισίου ανάλυσης $T=50\text{msec}$ και $\text{Toverlap} = 25\text{msec}$.

Προτεινόμενα χαρακτηριστικά τα οποία θα εξαχθούν με τη χρήση του MIRtoolbox:

- 1) Auditory Roughness (mirroughness)
Στατιστικές μετρήσεις που θα υπολογίσετε: μέση τιμή και τυπική απόκλιση όλων των παραθύρων ανάλυσης, μέση τιμή (mean) του υπολογισμένου roughness με τιμή μικρότερη του median και μέση τιμή (mean) του roughness με τιμή μεγαλύτερη του median
- 2) Rythmic Periodicity Along Auditory Channels (mirfluctuation)
Στατιστικές μετρήσεις: max of summarized fluctuation, mean of summarized fluctuation
- 3) Key Clarity (mirkey)
Στατιστικές μετρήσεις: μέση τιμή όλων των παραθύρων ανάλυσης
- 4) Modality (mirmode)
Στατιστικές μετρήσεις: μέση τιμή όλων των παραθύρων ανάλυσης
- 5) Spectral Novelty (mirsnovelty)
Στατιστικές μετρήσεις: μέση τιμή όλων των παραθύρων ανάλυσης
- 6) Harmonic Change Detection Function (HCDF) (mirhcdf)
Στατιστικές μετρήσεις: μέση τιμή όλων των παραθύρων ανάλυσης

Αποτέλεσμα της επεξεργασίας αυτής θα είναι 10 χαρακτηριστικά για κάθε μουσικό απόσπασμα.

Σας δίνετε επίσης η δυνατότητα να πειραματιστείτε με την συνάρτηση mirfeatures αλλά και να συμβουλευτείτε το manual1.6.1 του MIRtoolbox (σελ. 193) για άλλα πιθανά χαρακτηριστικά τα οποία βοηθάνε στην πρόβλεψη του συναισθήματος και συγκεκριμένα χαρακτηρίζουν διαφορετικές δομές της μουσικής, όπως η δυναμική, η χροιά, η αρμονία, ο ρυθμός κ.ά.

Βήμα 7

Εξαγωγή των Mel Frequency Cepstral Coefficients (MFCCs).

Ενδεικτικές τιμές για την εξαγωγή:

- Διάρκεια πλαισίου $T=25\text{msec}$ και η επικάλυψη $\text{Toverlap}=10\text{msec}$
- Αριθμός συστοιχίας φίλτρων $Q=26$
- Αριθμός ζητούμενων χαρακτηριστικών $N_c = 13$

Επίσης, εξάγετε την πρώτη και δεύτερη παράγωγο των MFCCs (σύνολο χαρακτηριστικών 39).

Οι στατιστικές μετρήσεις που θα υπολογίσετε μετά την εξαγωγή των MFCCs είναι:

Μέση τιμή, τυπική απόκλιση, μέση τιμή του 10% των μεγαλύτερων τιμών και μέση τιμή του 10% των μικρότερων τιμών για καθένα από τα 39 χαρακτηριστικά (MFCC+deltas+deltasdeltas).

Βήμα 8

Κατά την παρουσία σας στο 3^ο Εργαστήριο θα χρειαστείτε τους αλγορίθμους ταξινόμησης που υλοποιήσατε κατά τη διάρκεια της 1^{ης} Εργαστηριακής Άσκησης. Προετοιμάστε τους κώδικες που υλοποιήσατε για χρήση κατά την διάρκεια του 3^{ου} Εργαστηρίου.

Βήμα 9

Εγκαταστήστε και εξοικειωθείτε με το Weka, η χρήση το οποίου θα γίνει στο επόμενο μέρος της άσκησης.

ΠΑΡΑΔΟΤΕΑ

(1) Αναφορά (σε pdf) που θα περιγράφει τη διαδικασία που ακολουθήθηκε σε κάθε βήμα, καθώς και τα σχετικά αποτελέσματα.

(2) Κώδικας Matlab (συνοδευόμενος από σύντομα σχόλια).

Συγκεντρώστε τα (1) και (2) σε ένα .zip αρχείο το οποίο πρέπει να αποσταλεί μέσω του mycourses.ntua.gr εντός της καθορισμένης προθεσμίας.

Εργαλεία που θα χρησιμοποιηθούν στο εργαστήριο:

1) Matlab

2) MIRtoolbox

<https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox>

Για την εξαγωγή χαρακτηριστικών θα χρησιμοποιήσετε το MIRtoolbox, ένα open source εργαλείο το οποίο και δίνει την δυνατότητα εξαγωγής πολλών μουσικών χαρακτηριστικών. Στο zip που θα κατεβάσετε θα βρείτε αναλυτικά manuals με περιγραφές τόσο των διαφορετικών χαρακτηριστικών όσο και για τις εντολές που θα χρειαστείτε τελικά στο περιβάλλον του Matlab για την εξαγωγή των χαρακτηριστικών αλλά και τον υπολογισμό των στατιστικών τους.

3) Weka <http://www.cs.waikato.ac.nz/ml/weka/>

Το εργαλείο θα χρησιμοποιηθεί εν μέρει για την διαδικασία ταξινόμησης, 2^ο μέρος της Εργαστηριακής Άσκησης, και αποτελείται από διάφορους αλγορίθμους μηχανικής μάθησης.