

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ **ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

ΡΟΗ Σ

9ο Εξάμηνο

“Αναγνώριση Προτύπων”

3η Εργαστηριακής Άσκησης:

Αναγνώριση Συναισθήματος στη Μουσική

ΜΑΡΟΥΓΚΑΣ ΙΣΙΔΩΡΟΣ 03113048

ΜΠΑΚΟΣ ΠΑΝΑΓΙΩΤΗΣ 03113014

Ακαδημαϊκό Έτος 2017-2018

Εισαγωγή

Σκοπός της άσκησης αυτή είναι η υλοποίηση ενός συστήματος αυτόματης ταξινόμησης μουσικών αποσπασμάτων βάσει του συναισθήματος που το απόσπασμα αυτό επιδιώκει να προκαλέσει στο δέκτη.

Στο πλαίσιο αυτό, το συναίσθημα αναπαριστάται με τη χρήση 2 χαρακτηριστικών σε raw μορφή με τα μεγέθη Χαρά/Λύπη (Valence) και Ενεργοποίηση-/Απενεργοποίηση (Activation) κι αυτό καθώς οφείλονται σε διεγέρσεις διαφορετικών περιοχών νευρώνων του εγκεφάλου.

Στη συνέχεια για κάθε μια από τις 2 αυτές διαστάσεις βγάζουμε features 2 διαφορετικών κατηγοριών με τη χρήση του MIR ToolBox: features που αφορούν τη μουσική του εκάστοτε κομματιού (metrics που πηγάζουν από roughness, fluctuation, modality κλπ) και features για τη φωνή του τραγουδιστή και τα χρησιμοποιούμε για supervised training και στη συνέχεια classification-testing ενώ στο τέλος αυτά τα 2 συνδυάζονται με την ελπίδα καλύτερων αποτελεσμάτων.

Τα αποτελέσματα αυτά με τη βοήθεια μετρικών επίδοσης αλγορίθμων ταξινόμησης: accuracy, precision, recall, f1-score με μεγαλύτερη προσοχή να δίνεται στην πρώτη και την τελευταία αντί για τη διαισθητικά άμεση επιλογή μόνο της accuracy για μεγαλύτερη επισκόπηση της ποιότητας των αποτελεσμάτων και υψηλότερο robustness.

Πριν ξεκινήσουμε την εφαρμογή μοντέλων μάθησης παρατηρούμε από τα δεδομένα του 3ου Προεργαστηρίου πως ο 3ος Labeler έχει μεγάλη standard deviation κάτι που μεταφράζεται σε ευμετάβλητη άποψη για τα κομμάτια στις βαθμολογήσεις του σε αντίθεση με τους άλλους 2 και πως ο 1ος labeler εμφανίζει ισχυρό bias προς τα υψηλά Valence ενώ ο 3ος προς τα χαμηλά Activation.

Εν γένει υπάρχει μια ροπή των labelers προς τα υψηλά Valence.

Εκτέλεση

Βήμα 10

Για τη διαδικασία της ταξινόμησης θεωρούμε πως έχουμε ένα πρόβλημα όπου τα δείγματα με τιμές κάτω του 3 ανήκουν στην κατηγορία -1 (*High Activation* ή *Negative Valence*) και τα δείγματα με τιμές άνω του 3 ανήκουν στην κατηγορία 1 (*Low Activation* ή *Positive Valence*). Τα δείγματα με τιμή ίση με 3 στη μία τουλάχιστον από τις 2 διαστάσεις αγνοούνται. Προκύπτουν επομένως 355 δείγματα με Activation διάφορο του 3, και 359 δείγματα με Valence διάφορο του 3.

Κάνουμε ακόμα κανονικοποίηση της κατανομής (standarization) των δεδομένων, ώστε κάθε χαρακτηριστικό να αποκτήσει μέση τιμή ίση με 0 και τυπική απόκλιση ίση με 1. Με τον τρόπο αυτό παρατηρήσαμε αξιοσημείωτη βελτίωση στα αποτελέσματα

Βήμα 11

Δημιουργούμε 3 σύνολα χαρακτηριστικών :

- 1) Χαρακτηριστικά που χαρακτηρίζουν διαφορετικές δομές της μουσικής, όπως η δυναμική, η χροιά, η αρμονία, ο ρυθμός κτλ (τα οποία εξάγαμε στο βήμα 6)
- 2) Χαρακτηριστικά που αφορούν την ανθρώπινη φωνή (Mel Frequency Cepstrum Coefficients) (τα οποία εξάγαμε στο βήμα 7)
- 3) Συνδυασμός των 1) και 2)

Τα δεδομένα τα χωρίζουμε τυχαία σε ποσοστό 80% train set και 20% test set και χρησιμοποιούμε 3-fold cross-validation για να εξασφαλίσουμε πιο ακριβή υπολογισμό του ποσοστού επιτυχίας.

Βήμα 12

Αρχικά ταξινομούμε τα δεδομένα εφαρμόζοντας τον αλγόριθμο Nearest Neighbor που υλοποιήσαμε στο 1ο εργαστήριο για 1,3,5,7,11,13 γείτονες με τα αποτελέσματα των μετρικών accuracy και f1-score να παρουσιάζονται παρακάτω.

(Σημειώνεται εδώ και για παρακάτω ότι το άνω όριο για το f1-score όπως μας δόθηκε στην εκφώνηση είναι το 0.5 και όχι το 1. Για μια τέτοια περίπτωση ο αριθμητής του θα έπρεπε να πολλαπλασιάζεται επι 2 στον τύπο του).

Activation(set 1)		
7 γείτονες	Accuracy	0.7762
	F1-Score	0.3456

Valence (set 1)		
5 γείτονες	Accuracy	0.7136
	F1-Score	0.3184

Activation(set 2)		
11 γείτονες	Accuracy	0.7952
	F1-Score	0.3687

Valence (set 2)		
7 γείτονες	Accuracy	0.8357
	F1-Score	0.4582

Activation(set 3)		
7 γείτονες	Accuracy	0.8048
	F1-Score	0.3704

Valence (set 3)		
7 γείτονες	Accuracy	0.8310
	F1-Score	0.4907

Βήμα 13

Τώρα ταξινομούμε τα δεδομένα μας με τη χρήση του Naive Bayes Classifier που δημιουργήσαμε στο 1ο εργαστήριο. Η accuracy με βάση τις a-priori πιθανότητες είναι 55-45% για το Activation και 63%-37% για το Valence, οπότε αυτό θα είναι και το πρώτο διαπισθητικό κατώφλι απόδοσης που θα πρέπει να ξεπεράσουμε. “Εκπαιδεύουμε” τον ταξινομητή υπολογίζοντας τον μέσο όρο και τη διασπορά των features στο train set και προκρίνονται μετρικές αξιολόγησης :

Activation(set 1)	
Accuracy	0.7476
F1-Score	0.3575

Valence (set 1)	
Accuracy	0.6761
F1-Score	0.4096

Activation(set 2)	
Accuracy	0.7762
F1-Score	0.4073

Valence (set 2)	
Accuracy	0.6948
F1-Score	0.3939

Activation(set 3)	
Accuracy	0.8190
F1-Score	0.4481

Valence (set 3)	
Accuracy	0.7136
F1-Score	0.4024

Βήμα 14

Στη συνέχεια εφαρμόζουμε την τεχνική μείωσης της διαστατικότητας με τη μέθοδο Ανάλυσης σε Κύριες Συνιστώσες. Παρατηρούμε πως εν γένει τα προβλεπόμενα αποτελέσματα παρά τη μείωση της διαστατικότητας παραμένουν στα ίδια καλά επίπεδα για accuracy και f1score, πράγμα που μας κάνει να κρίνουμε τη μέθοδο ως επιτυχή. Συγκεκριμένα, παίζοντας με τυχαίους αύξοντες αριθμούς συνιστωσών πχ 4,18,25,40,55 παρατηρούμε ότι καθώς αυξάνουμε τις συνιστώσες οι μετρικές μας βελτιώνονται μέχρι τις 25 συνιστώσες οπότε και αρχίζουν να πέφτουν καθώς η curse of dimensionality(ειδικά στον Naive Bayes όπου το φαινόμενο αυτό προσθέρει model selection errors) επιδρά πλέον περισσότερο απότι το περιεχόμενο πληροφορίας των extra συνιστωσών.

Για τον kNN:(για τους γείτονες που έδωσαν και τα καλύτερα αποτελέσματα πριν ανά σετ δεδομένων)

Activation						
		Number of Components				
		4	18	25	40	55
7 γείτονες	Accuracy	0.7423	0.7912	0.7952	0.6751	0.6942
	F1-Score	0.3965	0.4216	0.4087	0.3754	0.3988

Valence						
		Number of Components				
		4	18	25	40	55
7 γείτονες	Accuracy	0.6818	0.7649	0.8015	0.7465	0.6995
	F1-Score	0.3632	0.4123	0.3923	0.3758	0.3655

Για τον Naive Bayes:

Activation					
	Number of Components				
	4	18	25	40	55
Accuracy	0.7524	0.7952	0.7952	0.6857	0.7619
F1-Score	0.3856	0.4146	0.4178	0.3737	0.4005

Valence					
	Number of Components				
	4	18	25	40	55
Accuracy	0.6948	0.7699	0.8122	0.7465	0.6995
F1-Score	0.3625	0.4033	0.3933	0.3768	0.3655

Βήμα 15

Φέρνουμε τα σύνολα δεδομένων που δημιουργήσαμε στο **Βήμα 10** στο κατάλληλο για το Weka format. Δημιουργούμε για κάθε ένα από τα 6 σύνολα χαρακτηριστικών ένα *.arff* αρχείο, στη μορφή που υποδεικνύει η εκφώνηση. Η συγκεκριμένη διαδικασία γίνεται στη συνάρτηση **WekaDataPrep.m**. Με την ολοκλήρωση του βήματος αυτού τελειώνει ο κώδικας του Matlab και πλέον εργαζόμαστε με τα 6 αυτά αρχεία στο Weka.

Βήμα 16

Στο βήμα αυτό πειραματιζόμαστε με αλγορίθμους ταξινόμησης του Weka. Συγκεκριμένα:

1) Multi Layer Perceptron (2 hidden layers):

Activation(set 1)	
Accuracy	0.7521
F1-Score	0.3261

Valence (set 1)	
Accuracy	0.7354
F1-Score	0.3677

Activation(set 2)	
Accuracy	0.6966
F1-Score	0.3483

Valence (set 2)	
Accuracy	0.7622
F1-Score	0.3833

Activation(set 3)	
Accuracy	0.7183
F1-Score	0.359

Valence (set 3)	
Accuracy	0.7632
F1-Score	0.3842

2) Support Vector Machines :

Activation(set 1)	
Accuracy	0.7775
F1-Score	0.38885

Valence (set 1)	
Accuracy	0.7855
F1-Score	0.3923

Activation(set 2)	
Accuracy	0.7408
F1-Score	0.3705

Valence (set 2)	
Accuracy	0.8050
F1-Score	0.4231

Activation(set 3)	
Accuracy	0.7549
F1-Score	0.3775

Valence (set 3)	
Accuracy	0.7548
F1-Score	0.3654

3) Random Forest :

Activation(set 1)	
Accuracy	0.7775
F1-Score	0.3765

Valence (set 1)	
Accuracy	0.7549
F1-Score	0.3754

Activation(set 2)	
-------------------	--

Accuracy	0.7690
F1-Score	0.3845

Valence (set 2)	
Accuracy	0.7827
F1-Score	0.3865

Activation(set 3)	
Accuracy	0.8000
F1-Score	0.4000

Valence (set 3)	
Accuracy	0.7632
F1-Score	0.3765

Βήμα 17

Ο κύριος στόχος της επιλογής χαρακτηριστικών είναι η αναζήτηση ενός βέλτιστου υποσυνόλου χαρακτηριστικών που θα οδηγεί σε ανώτερη ποιότητα ταξινόμησης ελαττώνοντας παράλληλα το όποιο υπολογιστικό κόστος σε σχέση με την περίπτωση που θα διατηρούσε όλα τα features.

Οι 2 κύριες μέθοδοι που προκρίνονται στην βιβλιογραφία και χρησιμοποιούνται συνηθέστερα στην πράξη είναι η Wrapper και η Filtering.

Τα Wrappers είναι πρακτικά αλγόριθμοι αναζήτησης στον feature space οι οποίοι δημιουργούν υποσύνολα-συνδυασμούς των χαρακτηριστικών τρέχοντας έναν επιλεγμένο αλγόριθμο που σκοπεύουμε έπειτα να χρησιμοποιήσουμε σε όλο το train set πάνω στο κάθε υποσύνολο. Στη συνέχεια εκτιμάται μια μετρική απόδοσης με βάση την οποία αποφασίζεται αν το τρέχον υποσύνολο θα επεκταθεί και σε επομένα features ή

συνδυασμούς αυτών.

Τα Wrappers έχουν το μειονέκτημα της υπολογιστικής ακρίβειας και του κινδύνου του overfitting.

Τα Filters είναι παρόμοια με τα Wrappers σε ότι αφορά τον Αλγόριθμο αναζήτησης αλλά αντί για την εκτίμηση μιας μετρικής επί ενός μοντέλου απλώς φιλτράρουν το τρέχον υποσύνολο των χαρακτηριστικών και κρίνουν ή όχι την ανάγκη επέκτασης του (σε forward μορφές).

Φυσικά, πρόκειται για πιο quick and dirty μεθοδολογία.

Backward υπολοποιήσεις τρέχουν ανάποδα και κλαδεύουν κάθε φορά το τρέχον υποσύνολο χαρακτηριστικών ξεκινώντας από όλα τα διαθέσιμα μαζί,

Αλγόριθμοι αναζήτησης που εν γένει χρησιμοποιούνται και στις 2 μεθοδολογίες είναι Άπληστη Αναζήτηση ή Τεχνικές Ανάβασης Λόφου.

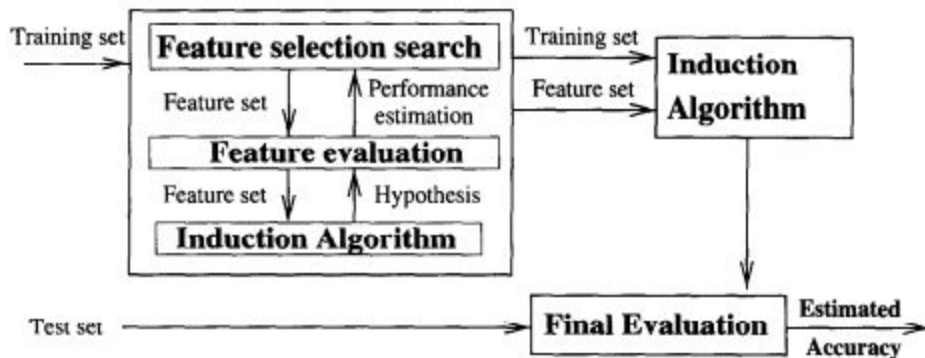


Fig. 1. The wrapper approach to feature subset selection. The induction algorithm is used as a "black box" by the subset selection algorithm.

Δοκιμάζουμε εκδοχές και των 2 μεθοδολογιών για διαφορετικούς συνδυασμούς Αλγορίθμων αναζήτησης και μετρικών αξιολόγησης για κάθε ταξινομητή σε κάθε διάσταση και αναγράφεται για τον καθένα ποια features χρησιμοποιήθηκαν.

Από άποψη απόδοσης η τεχνική Wrapper υπερισχύει (τα features επιλέχθηκαν μόνο πάνω στο SVM μοντέλο γιατί μόνο σε αυτό η πολυπλοκότητα υπολογισμών ήταν ανεκτή).

Για το συνδυαστικό Dataset 3:

Επιλέχθηκαν για το Activation:

modality hcdf_mean
mfcc_deltadelta_mean_7
mfcc_deltadelta_mean_11
mfcc_std_9
mfcc_delta_std_2
mfcc_deltadelta_std_2
mfcc_mean_low_2
mfcc_mean_low_13
mfcc_deltadelta_mean_high_7

Και για το Valence:

rough_mean2
mfcc_mean_1
mfcc_mean_2
mfcc_delta_mean_10
mfcc_deltadelta_mean_11
mfcc_std_2
mfcc_std_7
mfcc_std_10
mfcc_delta_std_4
mfcc_delta_std_10
mfcc_deltadelta_std_4
mfcc_deltadelta_std_10
mfcc_mean_low_1
mfcc_delta_mean_low_7
mfcc_mean_high_8

1) Multi Layer Perceptron (2 hidden layers):

Activation	
Accuracy	0.8000

F1-Score	0.4000
-----------------	--------

Valence	
Accuracy	0.7465
F1-Score	0.3733

2) Support Vector Machines :

Activation	
Accuracy	0.8507
F1-Score	0.4255

Valence	
Accuracy	0.8354
F1-Score	0.4095

3) Random Forest :

Activation	
Accuracy	0.7915
F1-Score	0.3958

Valence	
Accuracy	0.7688
F1-Score	0.3832