

# **Αναγνώριση Μουσικών Ειδών**

## **με Χρήση RBF Δικτύων**

Παπαϊωάννου Μαγδαληνή [7359]  
Σαββίδης Θεμιστοκλής [6078]

**Περιγραφή** -- Σκοπός της εργασίας αυτής είναι η μελέτη των νευρωνικών δικτύων ακτινικών συναρτήσεων (RBF) μέσα από το πρόβλημα της αυτοματοποιημένης κατηγοριοποίησης μουσικών κομματιών στα αντίστοιχα μουσικά τους είδη. Την περιγραφή του νευρωνικού δικτύου ακολουθεί υλοποίηση σε python, τα αποτελέσματα της οποίας αξιολογούνται τόσο με βάση τα αποτελέσματα παραπλήσιων υλοποιήσεων όσο και με την ανθρώπινη ικανότητα αναγνώρισης του είδους ενός μουσικού κομματιού.

Λέξεις - Κλειδιά: Νευρωνικό δίκτυο, ακτινική συνάρτηση, μουσικό είδος

### 1. Εισαγωγή

Η αναγνώριση του μουσικού είδους ενός μουσικού κομματιού είναι μία διαδικασία ιδιαίτερης πολυπλοκότητας που καθιστά την αυτοματοποίηση της εξαιρετικά δύσκολη. Η δυσκολία αυτή έγκειται τόσο στην υποκειμενικότητα της αντιληπτικής ικανότητας του ακροατή (ψυχοακουστική) όσο και στην αδυναμία της επιστημονικής κοινότητας να ποσοτικοποιήσει με ακρίβεια και συνέπεια χαρακτηριστικά του ηχητικού σήματος που να είναι ικανά να το περιγράψουν με κάποιο βαθύτερο, μη τετριμμένο τρόπο.

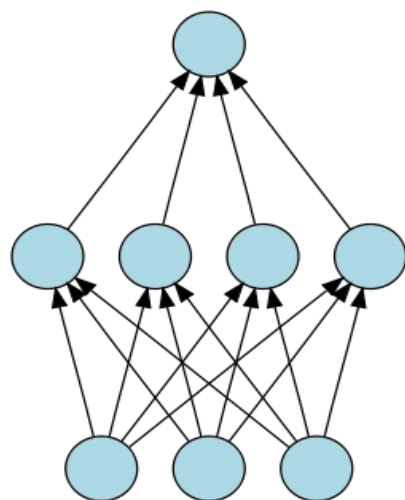
Η ιδιότητα των νευρωνικών δικτύων να εκπαιδεύονται με βάση τα χαρακτηριστικά παραδειγμάτων για τα οποία γνωρίζουμε ήδη την επιθυμητή έξοδο και να συμπεριφέρονται ως καθολικοί ταξινομητές, τα καθιστά ιδιαίτερα ελκυστικά μπροστά στο πρόβλημα αναγνώρισης του μουσικού είδους. Συγκεκριμένα, επιλέγεται ο τύπος δικτύων ακτινικών συναρτήσεων για αυτο-αναφορικούς και μη λόγους. Στους πρώτους συγκαταλέγονται ο σκοπός της άσκησης, η εύκολη πρόσβαση σε σχετική ερευνα και η προτροπή του κ. Δερματά. Στους δεύτερους και πιο ουσιαστικούς, η δυνατότητα χρήσης αλγορίθμων χωρίς επίβλεψη για την αρχικοποίηση των παραμέτρων της

ακτινικής συνάρτησης, η ιδιότητα τους να εκπαιδεύονται σε σχετικά μικρό χρονικό διάστημα με την προϋπόθεση ύπαρξης καλών αρχικών παραμέτρων και η ντετερμινιστική εύρεση των βαρών που βασίζεται στη χρήση ψευδοαντίστροφου πίνακα. Τέλος, η μόνη αυθαίρετη παράμετρος για την οποία έπρεπε εμείς να αποφασίσουμε ήταν ο αριθμός των νευρώνων του κρυφού επιπέδου.

Στην ενότητα 2 περιγράφουμε τους μηχανισμούς λειτουργίας ενός RBF δικτύου. Ακολουθεί η ενότητα 3 με την επεξήγηση των χαρακτηριστικών που χρησιμοποιήθηκαν και μια συνοπτική ματιά στα δύο μέρη που απαρτίζουν την υλοποίηση: αυτό της κατασκευής του συνόλου εκπαίδευσης και αυτό που μοντελοποιεί το δίκτυο. Τέλος, στην ενότητα 4 τα αποτελέσματα μας συγκρίνονται με αποτελέσματα άλλων ερευνητών, προτείνονται βελτιώσεις και απαριθμούνται κάποιες από τις παρατηρήσεις μας.

## 2. Ανατομία ενός RBF δικτύου

Τα RBF δίκτυα απαρτίζονται από ένα κρυφό επίπεδο και ένα επίπεδο εξόδου:



Output  $y$

Linear weights

Radial basis functions

Weights

Input  $x$

Πριν την είσοδο τους στο κρυφό επίπεδο, τα χαρακτηριστικά κάθε παραδείγματος σταθμίζονται από βάρη. Τα βάρη αυτά είτε προσαρμόζονται κατά τη διάρκεια της εκπαίδευσης είτε τίθενται ίσα με τη μονάδα, μέθοδος που ακολουθήθηκε και στη δική μας περίπτωση.

Η συνάρτηση ενεργοποίησης του κρυφού επιπέδου είναι μια ακτινική

συνάρτηση η οποία δίνει και το όνομα της σ' αυτόν τον τύπο δικτύου.

Ακτινικές συναρτήσεις ονομάζονται οι συναρτήσεις με ανεξάρτητη μεταβλητή την ευκλείδεια απόσταση του κάθε στοιχείου-παραδείγματος από ένα κέντρο. Στην υλοποίηση μας χρησιμοποιούμε την ακόλουθη:

$$\varphi_j(x) = \exp\left\{-\frac{\|x - \mu_j\|^2}{2\sigma_j^2}\right\}$$

όπου  $j$  ο  $j$ -οστός νευρώνας του κρυφού επιπέδου,  $x$  το διάνυσμα με τα χαρακτηριστικά του παραδείγματος,  $\mu_j$  το κέντρο του αντίστοιχου νευρώνα και  $\sigma_j$  η τυπική απόκλιση των αποστάσεων όλων των παραδειγμάτων που αντιστοιχούν στον νευρώνα  $j$  από το κέντρο του.

Τα κέντρα των ακτινικών συναρτήσεων μπορούν να βρεθούν αρκετά αποδοτικά με χρήση διάφορων αλγορίθμων ομαδοποίησης. Ένας εκ των επικρατέστερων είναι ο αλγόριθμος K-means, αφού είναι ένας γρήγορος, εξονυχτικά μελετημένος αλγόριθμος που συγκλίνει πάντα. Πρόκειται βέβαια για στοχαστικό αλγόριθμο το αποτέλεσμα του οποίου εξαρτάται από τα αρχικά αυθαίρετα επιλεγμένα κέντρα, κάτι που αποτελεί μειονέκτημα αλλά δεν καταφέρνει να επισκιάσει τα πλεονεκτήματά του. Έχουν προταθεί διάφοροι τρόποι ντετερμινιστικής αρχικοποίησης των κέντρων ώστε ο k-means να συγκλίνει σε αποτελεσματικά τοπικά ελάχιστα, όπως για παράδειγμα ο Furthest First και ο Subset Furthest First. Για λόγους απλότητας όμως, στην εργασία μας δεν ασχοληθήκαμε με αυτούς.

Στο πρώτο επίπεδο προσθέτουμε έναν επιπλέον νευρώνα με συνάρτηση μεταφοράς 1, σκοπός του οποίου είναι η μετατόπιση του γραμμικού αποτελέσματος της εξόδου ως προς τον οριζόντιο άξονα. Ονομάζεται bias και αποδεικνύεται ιδιαίτερα χρήσιμο σε μικρές αλλά ζωτικής σημασίας ρυθμίσεις της συμπεριφοράς του νευρωνικού δικτύου.

Το επίπεδο εξόδου αποτελείται από νευρώνες ίσους με τις επιθυμητές κατηγορίες διάκρισης, οι συναρτήσεις μεταφοράς των οποίων είναι γραμμικές:

$$y_k(x) = \sum_{j=1}^M w_{kj} \varphi_j(x) + w_{k0}$$

όπου  $y_k(x)$  η  $k$ -οστή έξοδος του δικτύου για το παράδειγμα  $x$ ,  $M$  το σύνολο των μονάδων του κρυφού επιπέδου,  $w_{kj}$  το βάρος που δίνει ο κρυφός νευρώνας  $j$  στο νευρώνα εξόδου  $k$  και  $\varphi_j(x)$  η συνάρτηση ενεργοποίησης του  $j$ -οστού νευρώνα στο κρυφό επίπεδο.

Συνεπώς, υπολογίζεται το άθροισμα των αποτελεσμάτων των συναρτήσεων του κρυφού επιπέδου, αφού όμως έχει πολλαπλασιαστεί πρώτα το καθένα με κάποιο βάρος. Γνωρίζοντας την έξοδο του κρυφού επιπέδου και τις επιθυμητές εξόδους του δικτύου για το σύνολο εκπαίδευσης, μπορούμε να υπολογίσουμε ντετερμινιστικά τα βάρη αυτά τα οποία ελαχιστοποιούν το σφάλμα εξόδου.

$$\text{Σφάλμα} = (b - Gw)^T (b - Gw)$$

όπου  $b$  η επιθυμητή έξοδος,  $w$  τα άγνωστα βάρη και  $G$  η έξοδος του κρυφού επιπέδου.

Έτσι προκύπτει:

$$\frac{\partial \text{Σφάλμα}}{\partial w} = 0 \Rightarrow G^T Gw = G^T b \Rightarrow w = (G^T G)^{-1} G^T b \Rightarrow w = \text{pinv}(G)b$$

### 3. Υλοποίηση

#### 3.1 Σύνολο εκπαίδευσης

Αρχικά σκεφτήκαμε να εξάγουμε εμείς οι ίδιοι τα χαρακτηριστικά των μουσικών κομματιών γράφοντας κάποιο κατάλληλο πρόγραμμα. Διαπιστώσαμε έγκαιρα πως κάτι τέτοιο είναι εκτός του σκοπού του μαθήματος και της εργασίας και στραφήκαμε σε εναλλακτικές λύσεις. Στην αναζήτηση μας εντοπίσαμε το Million Song Dataset που περιέχει χαρακτηριστικά για, κυριολεκτικά, 1.000.000 τραγούδια. Το μέγεθος ολόκληρου του αρχείου ήταν απαγορευτικό για κατέβασμα και το δειγματικό αρχείο με χαρακτηριστικά 1.000 κομματιών περιείχε υλικό από διάφορους, τυχαίους καλλιτέχνες χωρίς να υπάρχει κάποιο μεταδεδομένο για το είδος της μουσικής. Στην περιγραφή του συνόλου, οι δημιουργοί του ανέφεραν την κύρια πηγή από την οποία άντλησαν τα δεδομένα, το EchoNest (<http://www.echonest.com>), υπηρεσία που χρησιμοποιεί τόσο γνωστούς όσο και ιδιόκτητους αλγόριθμους για την εξαγωγή .

Χρησιμοποιώντας το API του EchoNest, καταφέραμε να συλλέξουμε εύκολα χαρακτηριστικά για περίπου 350 κομμάτια κλασσικής, metal και jazz μουσικής: loudness, loudness\_max, speechiness, timbre, danceability και energy. Κάποια από αυτά ήταν μεμονωμένες τιμές (π.χ. speechiness) ενώ για άλλα υπολογίσαμε τη διασπορά και τη μέση τιμή από ένα σύνολο τιμών που εμφανίζονται καθ'όλη τη διάρκεια του κομματιού.

Θέλοντας να δούμε τον τρόπο σκέψης και άλλων μελετητών στο θέμα, αναζητήσαμε σχετικά papers και τελικά το αποτέλεσμα της εργασίας μας είναι μια απλοποιημένη εκδοχή του μοντέλου που περιγράφεται στο *Fast Recognition of Musical Genres Using RBF Networks* των Douglas Turnbull και Charles Elkan.

Για την υλοποίηση του RBF δικτύου επιλέχθηκε η γλώσσα Python τόσο λόγω της πληθώρας επιστημονικών βιβλιοθηκών που έχουν γραφτεί γι' αυτή όσο και λόγω της απλής της σύνταξης που την καθιστά ιδιαίτερα ελκυστική στην ανάπτυξη μικρών, proof-of-concept προγραμμάτων αλλά και στην επιλογή της για production κώδικα. Βιβλιοθήκες για στατιστική, αναγνώρισης προτύπων, διανυσματικές πράξεις και αναπαράστασης δεδομένων όπως τα numpy, scipy και matplotlib αποδείχθηκαν εξαιρετικά χρήσιμα και φιλικά για την ανάπτυξη του προγράμματος. Τονίζεται πως δε χρησιμοποιήθηκε κάποια βιβλιοθήκη νευρωνικών δικτύων (π.χ. Pybrain).

Η αντικειμενοστραφής προσέγγιση που ακολουθήσαμε παρέχει μία φιλική επαφή στο χρήστη της βιβλιοθήκης μας:

```
[0] rbfnn = RBFNN(n_centroids=5)
[1] rbfnn.train(training_dataset_input, training_dataset_output)
[2] results = rbfnn.predict(test_dataset_input)
[3] print util.evaluate(test_dataset_output, results)
```

#### 4. Αξιολόγηση και συμπεράσματα

Οι Tzanetakis και Cook έχουν καταλήξει πειραματικά πως η ικανότητα του μέσου ανθρώπου να κατηγοριοποιήσει "σωστά" ένα κομμάτι μουσικής σε ένα από 10 είδη κυμαίνεται στο 70% και με χρήση support vector machines έχουν καταλήξει σε αποτελέσματα ίδιας ακρίβειας (71%). Σε παρόμοια κλίμακα κυμαίνονται και τα αποτελέσματα του RBF δικτύου των Turnbull και Elkan οι οποίοι υποστηρίζουν πως το ποσοστό αυτό φράσσεται όχι από την αδυναμία του δικτύου αλλά από την ανεπάρκεια των αλγορίθμων εξαγωγής χαρακτηριστικών.

Η ακρίβεια της δικής μας υλοποίησης, με κομμάτια από 2 μουσικά είδη (metal, κλασσική) και μόλις 4 κέντρα που προκύπτουν από k-means με τυχαία αρχικοποίηση, δίνει

ακρίβεια της τάξης του 67% (βάσει της μεθόδου U) ως 85% (βάσει της μεθόδου C) . Με πρόσθεση τρίτου είδους (jazz) και 5 κέντρα η ακρίβεια πέφτει στο 45% (βάσει της μεθόδου U) ως 55% (βάσει της μεθόδου C).

Τα ποσοστά αυτά θα μπορούσαν να βελτιωθούν δοκιμάζοντας διάφορες ακτινικές συναρτήσεις και επιλέγοντας αυτές με την πιο επιθυμητή συμπεριφορά ή να βελτιώνοντας τις παραμέτρους των ακτινικών συναρτήσεων χρησιμοποιώντας τον gradient descent στα αποτελέσματα κάθε epoch μέσω κάποιας υβριδικής μοντελοποίησης. Σημαντική διαφορά θα μπορούσε να επιτευχθεί επίσης μέσω στοχευμένης αρχικοποίησης του Kmeans με χρήση αλγορίθμων όπως ο Furthest First και ο Subset Furthest First που αναφέρθηκαν παραπάνω. Θα μπορούσαμε επίσης να επιλέξουμε προσεκτικότερα τα χαρακτηριστικά των τραγουδιών με βάση τα οποία γίνεται η εκπαίδευση.

Ο κώδικας για τη δημιουργία του dataset και την υλοποίηση του RBF δικτύου είναι διαθέσιμος στα github account μας:

<https://github.com/mel--mel/>

<https://github.com/sawidis>

## 5. Βιβλιογραφία

- *Αναγνώριση Προτύπων*, Ευάγγελος Δερματάς
- *Fast Recognition of Musical Genres Using RBF Networks*, Douglas Turnbull, Charles Elkan
- *Comparison between Traditional Neural Networks and Radial Basis Function Networks* , Tiantian Xie, Hao Yu and Bogdan Wilamowski
- [scipy.org](https://www.scipy.org/), [numpy.org](https://www.numpy.org/), [wikipedia](https://www.wikipedia.org/)