



ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΗΣ
ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΤΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Χρήση εργαλείων μηχανικής μάθησης και γλωσσικής
τεχνολογίας για την αναγνώριση εχθρικού λόγου σε
στίχους τραγουδιών

Υποψήφιος: ΦΩΤΙΟΣ ΦΩΤΗΣ, Π2016046

Επιβλέπουσα: ΚΑΤΙΑ ΛΗΔΑ ΚΕΡΜΑΝΙΔΟΥ

ΚΕΡΚΥΡΑ 2024

Χρήση Εργαλείων Μηχανικής Μάθησης Και Γλωσσικής Τεχνολογίας Για Την Αναγνώριση Εχθρικού Λόγου Σε Στίχους Τραγουδιών

Η πτυχιακή εργασία εγκρίθηκε από:

i) ΚΑΤΙΑ ΛΗΔΑ ΚΕΡΜΑΝΙΔΟΥ

ii) ΘΕΜΗΣ ΕΞΑΡΧΟΣ

iii) ΑΝΔΡΕΑΣ ΚΑΝΝΑΒΟΣ

ΠΕΡΙΕΧΟΜΕΝΑ

A' Εισαγωγή	4
A'.1 ΠΕΡΙΛΗΨΗ	4
A'.2 ΒΑΣΙΚΑ ΘΕΜΑΤΑ	5
A'.3 ΜΟΝΤΕΛΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ	6
A'.4 ΓΕΝΙΚΗ ΕΡΕΥΝΑ	7
 B' Παραδοσιακές Μέθοδοι	 18
B'.1 ΠΕΡΙΓΡΑΦΗ	18
B'.2 ΕΡΕΥΝΑ ΚΛΑΣΙΚΩΝ ΜΕΘΟΔΩΝ	20
 Γ' Μοντέλα Βαθιάς Μάθησης	 32
Γ'.1 ΘΕΜΑΤΟΛΟΓΙΑ	32
Γ'.2 ΕΡΕΥΝΑ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ	33
 Δ' Συμπεράσματα	 44
 ΒΙΒΛΙΟΓΡΑΦΙΑ	 47

ΚΕΦΑΛΑΙΟ Α΄

Α΄ Εισαγωγή

Α΄.1 ΠΕΡΙΛΗΨΗ

Η χρήση τεχνολογιών μηχανικής μάθησης και γλωσσικής τεχνολογίας για την ανίχνευση εχθρικού λόγου σε στίχους τραγουδιών αποτελεί έναν αναπτυσσόμενο τομέα έρευνας. Η κύρια ιδέα είναι να χρησιμοποιηθούν αλγόριθμοι που εκπαιδεύονται σε μεγάλα σύνολα δεδομένων για την αναγνώριση μοτίβων και χαρακτηριστικών που σχετίζονται με εχθρική ρητορική. Ο εχθρικός λόγος, που περιλαμβάνει τη ρητορική μίσους και την επιθετική γλώσσα, έχει σοβαρές επιπτώσεις στην κοινωνία. Ειδικότερα, η ανίχνευση αυτού του λόγου στους στίχους τραγουδιών είναι ιδιαίτερα σημαντική λόγω της ευρείας επιρροής της μουσικής.

Η ανασκόπηση εξετάζει πώς η μηχανική μάθηση και οι τεχνολογίες γλώσσας μπορούν να χρησιμοποιηθούν για την ανίχνευση εχθρικού λόγου. Η αναγνώριση τέτοιου περιεχομένου έχει καταστεί ακόμη πιο σημαντική λόγω της αυξανόμενης χρήσης του διαδικτύου και των κοινωνικών δικτύων. Η διαδικασία περιλαμβάνει την ανίχνευση και το φιλτράρισμα περιεχομένου που προάγει μίσος, βία ή διακρίσεις.

Η ανασκόπηση εστιάζει στην ανάγκη για αποτελεσματικά εργαλεία ανίχνευσης και πρόληψης, αξιοποιώντας τεχνικές υπολογιστικής επιστήμης όπως η μηχανική μάθηση για την ανάλυση και κατηγοριοποίηση του περιεχομένου. Επίσης, εξετάζονται προηγούμενες προσεγγίσεις, αλγόριθμοι, μέθοδοι και διαθέσιμοι πόροι στον τομέα της επιστήμης των υπολογιστών. Παρά τις προόδους, η ακριβής ανίχνευση εχθρικού λόγου παραμένει πρόκληση λόγω της έμμεσης φύσης της έκφρασής του, απαιτώντας συνεχή βελτίωση των αλγορίθμων.

Α'.2 ΒΑΣΙΚΑ ΘΕΜΑΤΑ

Ορισμός Εχθρικού Λόγου:

Ένα βασικό στοιχείο αυτής της ανασκόπησης είναι ο ορισμός του εχθρικού λόγου στο πλαίσιο των στίχων τραγουδιών. Ο εχθρικός λόγος μπορεί να περιλαμβάνει ρητορική μίσους, προσβλητική γλώσσα, βίαιες απειλές, μισογυνικό, ομοφοβικό, ρατσιστικό ή ξενοφοβικό περιεχόμενο. Η πρόκληση έγκειται στην αναγνώριση των αποχρώσεων και του πλαισίου μέσα στη μουσική, η οποία μπορεί να είναι μεταφορική, συμβολική ή συναισθηματικά φορτισμένη.

Μηχανική Μάθηση στην Ταξινόμηση Κειμένων:

Οι τεχνικές μηχανικής μάθησης, ειδικά οι εποπτευόμενες μέθοδοι όπως οι Μηχανές Διανυσμάτων Υποστήριξης (SVM), τα δέντρα αποφάσεων, καθώς και τα μοντέλα βαθιάς μάθησης (όπως τα CNNs και RNNs), παίζουν σημαντικό ρόλο στην ανίχνευση εχθρικού λόγου. Αυτές οι μέθοδοι εκπαιδεύονται σε εκτενή σύνολα επισημασμένων δεδομένων, όπως στίχοι ή δείγματα ομιλίας, και χρησιμοποιούνται για να προβλέψουν αν ένα νέο τραγούδι περιέχει βλαβερό περιεχόμενο. Μοντέλα ανάλυσης συναισθημάτων και εξειδικευμένα μοντέλα για την αναγνώριση ρητορικής μίσους, όπως το BERT, μπορούν επίσης να αξιοποιηθούν. Προεκπαιδευμένα μοντέλα σε μεγάλες βάσεις δεδομένων, όπως στίχοι τραγουδιών ή tweets, προσφέρουν προσαρμοστικότητα στην ανάλυση των στίχων.

Τεχνικές Επεξεργασίας Φυσικής Γλώσσας:

Τα εργαλεία NLP βοηθούν στην προεπεξεργασία, ανάλυση και εξαγωγή σημαντικών χαρακτηριστικών από στίχους τραγουδιών. Τεχνικές όπως η τοκενικοποίηση, η λεμματοποίηση και οι ενσωματώσεις λέξεων διευκολύνουν την αναγνώριση εχθρικού λόγου. Επιπλέον, η ανάλυση συναισθημάτων, η ανίχνευση συναισθημάτων και η μοντελοποίηση θεμάτων χρησιμοποιούνται για την ανίχνευση αρνητικού περιεχομένου στους στίχους.

Προκλήσεις στην Ανάλυση Στίχων Τραγουδιών:

Οι στίχοι τραγουδιών θέτουν μοναδικές προκλήσεις σε σύγκριση με άλλες μορφές κειμένου, όπως τα tweets ή τα άρθρα. Αυτές περιλαμβάνουν:

i) Δημιουργική Χρήση Γλώσσας:

Οι καλλιτέχνες χρησιμοποιούν συχνά μεταφορική γλώσσα, μεταφορές και αργκό, τα οποία είναι δύσκολα στην ερμηνεία από τα παραδοσιακά μοντέλα.

ii) Πολιτισμικό Πλαίσιο και Υποκειμενικότητα:

Οι στίχοι συχνά αντικατοπτρίζουν συγκεκριμένα πολιτισμικά ή κοινωνικά ζητήματα. Αυτό που μπορεί να θεωρηθεί εχθρικό σε έναν πολιτισμό, μπορεί να ερμηνευτεί διαφορετικά σε άλλον.

iii) Πολυτροπικές Πτυχές:

Τα τραγούδια συνδυάζουν τους στίχους με τη μουσική, και ο τόνος ή το μουσικό στυλ μπορεί να αλλάξει τον τρόπο που αντιλαμβάνονται οι στίχοι. Μια καθαρά κειμενική ανάλυση μπορεί να χάσει τις αποχρώσεις που συλλαμβάνονται από τη μουσική.

iv) Ασάφεια και Σαρκασμός:

Ο εχθρικός λόγος στους στίχους μπορεί να μην είναι άμεσος αλλά υπονοούμενος, σαρκαστικός ή κωδικοποιημένος με τρόπους που ξεφεύγουν από τα τυπικά μοντέλα.

Α'.3 ΜΟΝΤΕΛΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Τα Μοντέλα Μηχανικής Μάθησης είναι υπολογιστικά μοντέλα που επιτρέπουν σε έναν υπολογιστή να μαθαίνει από δεδομένα χωρίς να προγραμματίζεται ρητά για κάθε πιθανή λύση. Μέσω της εκπαίδευσης σε σύνολο δεδομένων, τα μοντέλα αυτά μπορούν να αναγνωρίζουν μοτίβα και να κάνουν προβλέψεις ή αποφάσεις. Συγκεκριμένα, τα μοντέλα εκπαιδεύονται με δεδομένα που περιέχουν ζεύγη εισόδου και εξόδου (ετικέτες). Στόχος είναι να μάθουν τη σχέση μεταξύ των εισόδων και των εξόδων, ώστε να μπορούν να προβλέψουν την έξοδο για νέα δεδομένα.

Μερικά δημοφιλή Μοντέλα Μηχανικής Μάθησης:

i) Γραμμική Παλινδρόμηση (Linear Regression):

Χρησιμοποιείται για την πρόβλεψη συνεχών τιμών, βασισμένη στη γραμμική σχέση μεταξύ της εισόδου και της εξόδου.

ii) Δέντρα Αποφάσεων (Decision Trees):

Χρησιμοποιούν ένα διαδοχικό σύστημα αποφάσεων με βάση τις ιδιότητες των δεδομένων, για να καταλήξουν σε πρόβλεψη ή απόφαση.

iii) Naive Bayes:

Ένα στατιστικό μοντέλο που βασίζεται στο Θεώρημα του Bayes, ιδανικό για ταξινόμηση κειμένου και ανίχνευση σπαμ.

iv) Μηχανές Υποστήριξης Διανυσμάτων (SVM):

Χρησιμοποιούν υπερεπίπεδα για να διαχωρίσουν τα δεδομένα σε κατηγορίες με τον καλύτερο δυνατό τρόπο.

v) Νευρωνικά Δίκτυα (Neural Networks):

Προσομοιώνουν τη λειτουργία των νευρώνων του εγκεφάλου και χρησιμοποιούνται για σύνθετες εργασίες όπως η αναγνώριση εικόνων και ομιλίας.

vi) Συστήματα Συστάσεων:

Προτείνουν αντικείμενα (π.χ. προϊόντα, τραγούδια) στους χρήστες με βάση τα ενδιαφέροντά τους ή παρόμοιες προτιμήσεις άλλων χρηστών.

Α'.4 ΓΕΝΙΚΗ ΕΡΕΥΝΑ

EXPLICIT CONTENT DETECTION IN MUSIC LYRICS USING MACHINE LEARNING

Στη συγκεκριμένη μελέτη, προτείνονται πολλά μοντέλα μηχανικής μάθησης που αυτόματα εντοπίζουν ρητά περιεχόμενα σε Κορεάτικους στίχους και συγκρίνονται με παραστάσεις. Το προτεινόμενο Bagging μοντέλο με επιλεκτικό λεξιλόγιο ξεπέρασε όχι μόνο τα άλλα ανταγωνιστικά μοντέλα αλλά και την μέθοδο φιλτραρίσματος που χρησιμοποιεί το ανθρωπογενές λεξικό βωμολοχιών, το οποίο είναι μια ευρέως χρησιμοποιούμενη μέθοδος ανίχνευσης ρητού περιεχομένου στον συγκεκριμένο κλάδο. Η προτεινόμενη αυτοματοποιημένη προσέγγιση προβολής στίχων συμβάλλει πρακτικά στη μουσική βιομηχανία, βοηθώντας την να εξοικονομήσει σημαντικά χρόνο και προσπάθεια για ακατάλληλο περιεχόμενο για τους νέους.

ΜΕΘΟΔΟΣ:

Η κύρια συνθήκη του πειράματός είναι η χρήση μιας μεθόδου ταξινόμησης τύπου "Fail" ή "Pass" για τον εντοπισμό ακατάλληλων περιεχομένων, όταν οι στίχοι των τραγουδιών περιέχουν λέξεις από ένα λεξικό βωμολοχίας που διατίθεται στο Namu-Wiki. Το Namu-Wiki είναι μια παραλλαγμένη έκδοση της Wikipedia, η οποία έχει αναπτυχθεί από Κορεάτες χρήστες του διαδικτύου και αποτελεί το πιο δημοφιλές "Wiki-site" στην Κορέα, ξεπερνώντας κατά πολύ σε επισκεψιμότητα την Κορεάτικη Wikipedia. Το λεξικό βωμολοχίας του NamuWiki είναι αποτέλεσμα συνεργασίας πολλών χρηστών με τη μέθοδο του "Crowdsourcing", σε αντίθεση με άλλα λεξικά. Λόγω της μεγάλης απήχησης του NamuWiki στην Κορέα, στη συγκεκριμένη έρευνα χρησιμοποιείται το λεξικό βωμολοχίας του NamuWiki.

ΔΙΑΝΥΣΜΑ BOOLEAN & IDF:

i) Διάνυσμα Boolean: Η συχνότητα της βωμολοχίας όπου εμφανίζεται σε στίχους τραγουδιών δεν είναι σημαντική. Ακόμη και αν η υβριστική λέξη εμφανίζεται μόνο μία φορά στους στίχους, το τραγούδι θα ταξινομηθεί ως αποτυχία "Fail". Ως εκ τούτου, χρησιμοποιήθηκαν

πληροφορίες σχετικά με το εάν μία ρητή λέξη εμφανίζεται στο έγγραφο (π.χ εμφανίζεται “1” , δεν εμφανίζεται “0”) ως χαρακτηριστικό αντί για τον όρο συχνότητα.

ii) Διάνυσμα IDF: Η αναλογία των δεδομένων μετάβασης προς την αποτυχία ήταν περίπου 28/1, τα δεδομένα αποτυχίας ήταν σχετικά μικρά. Επειδή οι ρητές λέξεις είναι πιθανό να εμφανίζονται μόνο στα δεδομένα αποτυχίας, οπότε αυτές οι λέξεις θα είχαν υψηλό σκορ IDF. Η μέση βαθμολογία IDF των λέξεων στο λεξικό βωμολογιών ήταν 8.80, το οποίο ήταν συγκριτικά αρκετά υψηλό λαμβάνοντας υπόψη ότι το IDF βαθμολογεί ολόκληρο το λεξιλόγιο από 1.38 μέχρι και 10.22.

ΜΟΝΤΕΛΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΓΙΑ ΤΑΞΙΝΟΜΗΣΗ ΡΗΤΟΥ ΠΕΡΙΕΧΟΜΕΝΟΥ:

Στην παρούσα μελέτη χρησιμοποιήθηκαν δύο γνωστοί αλγόριθμοι ταξινόμησης, οι Bagging και AdaBoost. Ο αλγόριθμος AdaBoost είναι γνωστός για την ικανότητά του να παρέχει ακριβή αποτελέσματα ταξινόμησης σε πολλές πειραματικές δοκιμές και έχει αποδειχθεί ιδιαίτερα αποτελεσματικός σε περιπτώσεις ταξινόμησης κειμένου. Από την άλλη πλευρά, το Bagging, μια από τις πιο απλές μορφές του συνδυαστικού ταξινομητή, λειτουργεί μέσω της επαναληπτικής εκπαίδευσης μοντέλων δυαδικής ταξινόμησης, χρησιμοποιώντας τυχαία δείγματα από το σύνολο δεδομένων εκπαίδευσης. Οι επιμέρους ταξινομητές που προκύπτουν δημιουργούν ένα σύνολο, το οποίο μπορεί να παράγει το τελικό μοντέλο ταξινόμησης με τη χρήση απλής συνάθροισης.

ΣΥΜΠΕΡΑΣΜΑ:

Μετά από εξέταση των αποτελεσμάτων, εφαρμόζοντας το “AdaBoost” και “Bagging” αλγόριθμους. Οι στίχοι μετατράπηκαν σε διανύσματα των οποίων η διάσταση είναι το μέγεθος των χρησιμοποιημένων δεδομένων του λεξιλογίου.

Στην συγκεκριμένη έρευνα έγινε σύγκριση σε τρεις μορφές του διανύσματος:

- i) Χρησιμοποιώντας ολόκληρο το λεξιλόγιο και η συχνότητα όρου (TF) ως τιμή του διανύσματος
- ii) Χρησιμοποιώντας ολόκληρο το λεξιλόγιο και τη βαθμολογία (IDF) ως τιμή του διανύσματος (αγνοώντας το TF)
- iii) Χρησιμοποιώντας επιλεκτικό λεξιλόγιο με “ετικέτες σε τμήματα της ομιλίας” και βαθμολογία (IDF) ως τιμή του διανύσματος.

Όλα τα μοντέλα υλοποιήθηκαν χρησιμοποιώντας το Scikit-Learn Package για Python. Χρησιμοποιήθηκε ένα δέντρο απόφασης ως βασικός εκτιμητής στο μοντέλο AdaBoost και ο αριθμός των εκτιμητών ήταν 50. Επίσης, δέντρο απόφασης χρησιμοποιήθηκε και στο μοντέλο Bagging ως βασικός εκτιμητής και ο αριθμός των εκτιμητών ήταν 40.

COMPARING AUTOMATED METHODS TO DETECT EXPLICIT CONTENT IN SONG LYRICS

Η Parental Advisory Label (PAL) είναι μια προειδοποιητική ετικέτα που τοποθετείται σε ηχογραφήσεις για να αναγνωρίσει τη χρήση βωμολοχιών ή ακατάλληλων αναφορών, με στόχο την ενημέρωση των γονέων σχετικά με περιεχόμενο που ενδέχεται να μην είναι κατάλληλο για παιδιά. Από το 2015, οι ψηφιακοί πάροχοι όπως το iTunes, το Spotify, το Amazon Music κ.α ακολουθούν επίσης τις οδηγίες της PAL και επισημαίνουν τέτοιες ηχογραφήσεις ως 'ρητές'. Σήμερα, αυτή η σήμανση γίνεται κυρίως χειροκίνητα και σε εθελοντική βάση, με τα μειονεκτήματα ότι είναι χρονοβόρα, δαπανηρή, επιρρεπής σε λάθη και εν μέρει υποκειμενική. Στην παρούσα εργασία συγκρίνονται αυτοματοποιημένες μέθοδοι, από αναζητήσεις βασισμένες σε λεξικά έως προηγμένα βαθιά νευρωνικά δίκτυα, για την αυτόματη ανίχνευση ρητών περιεχομένων σε αγγλικούς στίχους. Τα πιο περίπλοκα μοντέλα παρέχουν μόνο ελαφρώς καλύτερα αποτελέσματα και, δίνοντας βάση στη ποιοτική ανάλυση των δεδομένων, συζητείται η θεμελιώδη δυσκολία και υποκειμενικότητα της διαδικασίας.

ΜΕΘΟΔΟΣ:

Στη συγκεκριμένη εργασία, γίνεται σύγκριση διαφόρων μεθόδων ταξινόμησης για την ανίχνευση ρητών στίχων. Όλες οι μέθοδοι έχουν κοινό ότι κατατάσσουν ολόκληρο το τραγούδι σε μία από τις δύο κατηγορίες: ρητό ή μη ρητό. Αυτό σημαίνει ότι η απόφαση για το αν ένας στίχος είναι ρητός λαμβάνεται συνολικά. Εξετάζεται η απόδοση διαφορετικών μεθόδων ταξινόμησης, από απλές αναζητήσεις σε λεξικά και ελέγχους λεξιλογίου έως μοντέλα βαθιάς μάθησης για γενική κατανόηση της γλώσσας.

ΜΕΘΟΔΟΙ ΒΑΣΙΣΜΕΝΕΣ ΣΕ ΛΕΞΙΚΟ

Η πιο απλή προσέγγιση για την εφαρμογή μιας αυτοματοποιημένης μεθόδου ανίχνευσης ρητού περιεχομένου είναι η χρήση ενός λεξικού ρητών λέξεων. Το λεξικό μπορεί να είναι χειροποίητο ή να έχει δημιουργηθεί αυτόματα από παραδείγματα ρητών και καθαρών στίχων. Στη συνέχεια, ένας ταξινομητής χρησιμοποιεί αυτό το λεξικό για να προβλέψει την κατηγορία ενός άγνωστου κειμένου τραγουδιού.

ΔΗΜΙΟΥΡΓΙΑ ΛΕΞΙΚΟΥ

Έχει προηγουμένως αποδειχθεί ότι η απόδοση με τη χρήση ενός αυτόματα δημιουργημένου λεξικού είναι καλύτερη από αυτή του χειροποίητου δημιουργημένου λεξικού. Επομένως, εξετάζεται μόνο η περίπτωση του λεξικού που έχει δημιουργηθεί από μηχανή. Έγινε δημιουργία ενός λεξικού, με λέξεις που είναι ενδεικτικές για ρητούς στίχους. Ορίστηκε η σημασία μιας λέξης για ρητούς στίχους, συγκριτικά με τη συχνότητά της σε καθαρούς στίχους. Αφαιρέθηκαν μοναδικές και πολύ κοινές λέξεις και περιορίστηκε ο αριθμός των όρων σε 1.000 για να αποφευχθεί η υπερβολική εξάρτηση από όρους που είναι πολύ ειδικοί για το σώμα δεδομένων.

ΑΝΑΖΗΤΗΣΗ ΜΕ ΛΕΞΙΚΟ

Δεδομένου ενός λεξικού D_n , αυτή η μέθοδος απλά ελέγχει αν το κείμενο ενός τραγουδιού S περιέχει οποιονδήποτε από τους ρητούς όρους που ορίζονται στο D_n . Στη συνέχεια, το S ταξινομείται ως ρητό μόνο αν περιέχει τουλάχιστον έναν ρητό όρο από το D_n .

ΠΑΛΙΝΔΡΟΜΗΣΗ ΜΕ ΛΕΞΙΚΟ

Αυτή η μέθοδος χρησιμοποιεί το BOW (Bag of Words) που έχει δημιουργηθεί από το D_n ως σύνολο χαρακτηριστικών για έναν ταξινομητή. Χρησιμοποιήθηκε λογιστική παλινδρόμηση, αλλά έχουν χρησιμοποιηθεί και RF (Random Forest) ή SVM (Support Vector Machine) σε παρόμοιες εφαρμογές.

ΠΑΛΙΝΔΡΟΜΗΣΗ BOW ΜΕ TF-IDF

Παρόμοια με την Παλινδρόμηση με Λεξικό, αλλά το BOW περιέχει το σύνολο του λεξιλογίου ενός δείγματος εκπαίδευσης αντί μόνο των ρητών όρων. Τα χαρακτηριστικά των λέξεων ζυγίζονται με το γνωστό σύστημα βαρύτητας tf-idf.

ΜΕΤΑΣΧΗΜΑΤΙΣΤΕΣ

Αυτά τα μοντέλα είναι δομημένα ως encoder-decoder και εκπαιδεύονται σε μη επιβλεπόμενα καθήκοντα προκειμένου να μάθουν πυκνές αναπαραστάσεις προτάσεων ή εγγράφων. Αυτά τα μοντέλα διαφέρουν από τα πιο παραδοσιακά αναδρομικά νευρωνικά δίκτυα σε διάφορες πτυχές. Ειδικότερα, ενώ τα αναδρομικά μοντέλα μπορούν να επεξεργαστούν ακολουθίες με τη σειρά, οι μετασχηματιστές (transformers) χρησιμοποιούν ένα κοινό μοντέλο του δεξιού και αριστερού συμφραζόμενου κάθε λέξης προκειμένου να κωδικοποιήσουν μια ολόκληρη ακολουθία ή έγγραφο. Επιπλέον, οι μετασχηματιστές είναι συνήθως λιγότερο υπολογιστικά δαπανηροί από τα αναδρομικά μοντέλα. Ένα από τα πιο επιτυχημένα μοντέλα βασισμένα σε transformer που προτάθηκαν τα τελευταία χρόνια είναι το BERT.

TEXTUAL DECONVOLUTION SALIENCY

Χρησιμοποιείται το μοντέλο TDS, το οποίο είναι ένα Συνελικτικό Νευρωνικό Δίκτυο (CNN) για την ταξινόμηση κειμένων. Αυτό το μοντέλο είναι απλό και περιλαμβάνει μια επίπεδη στρώση ενσωμάτωσης για την αναπαράσταση των λέξεων, μια συνελικτική στρώση με μέγιστη συγκέντρωση και δύο πλήρως συνδεδεμένες στρώσεις. Το ιδιαίτερο χαρακτηριστικό του μοντέλου είναι η ικανότητά του να αναστρέφει τη συνελικτική διαδικασία. Χρησιμοποιώντας τον χάρτη χαρακτηριστικών που έχει μάθει, τον επαναλαμβάνουν για να δημιουργήσουν ένα δείγμα τριών διαστάσεων. Εφαρμόζεται αυτό το μοντέλο για να βρεθούν τοπικές εξηγήσεις που συνδέονται με την παγκόσμια απόφαση της ταξινόμησης ως ρητή ή μη ρητή.

ΣΥΜΠΕΡΑΣΜΑ:

Η ταξινόμηση των στίχων τραγουδιών ως ρητοί ή μη ρητοί είναι μια αναπόφευκτα δύσκολη εργασία, καθώς αυτό που θεωρείται προσβλητικό εξαρτάται σε μεγάλο βαθμό από πολιτισμικές πτυχές που μπορεί να αλλάζουν με την πάροδο του χρόνου. Τα επιφανειακά μοντέλα που βασίζονται αποκλειστικά σε λεξικό βωμολοχιών επιτυγχάνουν απόδοση συγκρίσιμη με αυτή των βαθιών νευρωνικών δικτύων. Επίσης, η χειροκίνητη σήμανση είναι εξαιρετικά υποκειμενική, καθιστώντας προβληματική την αυτόματη ανίχνευση αν ένα κείμενο τραγουδιού πρέπει να χαρακτηριστεί ως ρητό ή μη ρητό. Εάν γίνει παρουσίαση σε μια αρμόδια αρχή μια αναφορά για λέξεις-κλειδιά που βρέθηκαν, συμφραζόμενα σεξουαλικού περιεχομένου και παρόμοια, μπορεί να καταλήξουν στη δική τους υποκειμενική συμπερασματική απόφαση για την τελική ετικέτα του κειμένου.

EVALUATION OF FREQUENTLY USED AUDIO FEATURES FOR CLASSIFICATION OF MUSIC INTO PERCEPTUAL CATEGORIES

Η συνεχώς αυξανόμενη διαθεσιμότητα μουσικής οδηγεί σε μεγαλύτερη ζήτηση για εφαρμογές Ανάκτησης Πληροφοριών Μουσικής (MIR), όπως είναι οι εφαρμογές σύστασης μουσικής και οι αλγόριθμοι αυτόματης κατηγοριοποίησης. Όταν οι εφαρμογές αυτές βασίζονται σε ηχητικά δεδομένα, ένα βασικό στοιχείο τους είναι οι διαδικασίες εξαγωγής ηχητικών χαρακτηριστικών. Στην παρούσα εργασία, αξιολογείται η αποτελεσματικότητα διαφόρων συνδυασμών εξαγωγής χαρακτηριστικών και αλγορίθμων μηχανικής μάθησης στην ταξινόμηση της μουσικής σε αντιληπτές κατηγορίες. Οι κατηγορίες που εξετάζονται περιλαμβάνουν το αντιληπτό τέμπο, τη διάθεση (χαρούμενη, ουδέτερη, λυπημένη), το συναίσθημα (απαλό, ουδέτερο, επιθετικό), την πολυπλοκότητα και το φωνητικό περιεχόμενο. Στόχος είναι να εντοπιστούν ποια χαρακτηριστικά της μουσικής δεν καταγράφονται από τις συνηθισμένες ηχητικές περιγραφές.

ΣΥΛΛΟΓΗ ΤΡΑΓΟΥΔΙΩΝ ΚΑΙ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ:

Για τα πειράματα, χρησιμοποιήθηκε μια εσωτερική βάση δεδομένων που περιλαμβάνει 834 κομμάτια σε μορφή mp3. Τα κομμάτια αυτά επισημάνθηκαν χειροκίνητα σε κατηγορίες. Εκτός από τις κατηγορίες genre2 και focus, οι υπόλοιπες βασίζονται στη λογική: Για παράδειγμα, η πολυπλοκότητα μπορεί να εξαρτάται από το είδος του ρυθμού, τη μελωδία ή τον αριθμό των διαφορετικών οργάνων που συμμετέχουν.

Οι κατηγοριοποιήσεις επιλέχθηκαν για να αντικατοπτρίζουν σημαντικές πτυχές της μουσικής που δεν είναι απαραίτητα συσχετισμένες. Για ένα σύστημα MIR (π.χ. σύστημα σύστασης

μουσικής) θα ήταν πολύτιμο να μπορεί να διακρίνει αυτές τις κατηγορίες, καθώς αποτελούν σημαντικό μέρος του τρόπου με τον οποίο οι άνθρωποι αναφέρονται στη μουσική.

ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ:

Για την εξαγωγή χαρακτηριστικών από τη μουσική, κάθε κομμάτι μετατράπηκε σε μορφή wav, μειώθηκε σε 11025 kHz mono, και χρησιμοποιήθηκαν ακριβώς 30 δευτερόλεπτα από τη μέση του κομματιού για τον υπολογισμό των χαρακτηριστικών. Για κάποια από τα χαρακτηριστικά, το ηχητικό απόσπασμα χωρίστηκε σε πλαίσια μήκους 256 δειγμάτων με επικάλυψη κατά το ήμισυ.

ΑΛΓΟΡΙΘΜΟΙ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ:

Για να εκτιμηθεί η ποσότητα των πληροφοριών που περιέχονται στα χαρακτηριστικά, χρησιμοποιήθηκαν τα δεδομένα αυτών για να εκπαιδευτούν και να γίνει η εφαρμογή των αλγορίθμων μηχανικής μάθησης. Αυτοί οι αλγόριθμοι είναι σχεδιασμένοι να ανακαλύπτουν την υποκείμενη δομή στα δεδομένα και να εξάγουν χρήσιμες πληροφορίες για την εκάστοτε ταξινόμηση. Έτσι, το ποσοστό επιτυχίας τους αποτελεί ένδειξη της ποιότητας των χαρακτηριστικών. Η επιτυχία αυτή μπορεί να μετρηθεί από τη μέση ακρίβεια ταξινόμησης, η οποία θα πρέπει να είναι σημαντικά υψηλότερη από τη γραμμή βάσης, δηλαδή την ακρίβεια που επιτυγχάνεται όταν ανατίθεται πάντα η πιο συχνή κατηγορία.

Επειδή οι διάφοροι αλγόριθμοι μηχανικής μάθησης ακολουθούν διαφορετικές μεθόδους, δεν αποδίδουν το ίδιο αποτελεσματικά σε όλα τα είδη δεδομένων. Γι' αυτόν τον λόγο, χρησιμοποιήθηκαν αρκετές παραλλαγές αλγορίθμων μηχανικής μάθησης από το εργαλείο WEKA.

- i) K-Nearest Neighbors, with $K \in \{1, 3, 5, 10\}$
- ii) Naive Bayes, additional with a kernel estimation algorithm
- iii) The C4.5 algorithm (a decision tree learner)
- iv) A Support Vector Machine (SVM)
- v) AdaBoost with C4.5 and with Decision Stump
- vi) Classification via Regression, applying M5 and linear regression

Σε αυτή τη διαδικασία, οι αλγόριθμοι μηχανικής μάθησης εκπαιδεύτηκαν με το 90% των κομματιών, για τα οποία ήταν γνωστή η κατηγορία τους, και στη συνέχεια χρησιμοποιήθηκαν για να ταξινομήσουν το υπόλοιπο 10%. Η διαδικασία επαναλήφθηκε δέκα φορές, ώστε κάθε κομμάτι να ταξινομηθεί μία φορά.

ΣΥΜΠΕΡΑΣΜΑ:

Τα συνολικά αποτελέσματα ήταν ελάχιστα πάνω από τη γραμμή βάσης για τους περισσότερους συνδυασμούς συνόλων χαρακτηριστικών και αλγορίθμων εκμάθησης στις περισσότερες από τις εξεταζόμενες κατηγοριοποιήσεις. Σε όλες τις περιπτώσεις, οι καλύτερες τιμές ακρίβειας που επιτεύχθηκαν από τον αντίστοιχα προτεινόμενο αλγόριθμο, ήταν πιο καλές, από τις καλύτερες τιμές ακρίβειας που επιτεύχθηκαν από τα άλλα τρία σύνολα χαρακτηριστικών. Από τις εξεταζόμενες κατηγοριοποιήσεις, οι αλγόριθμοι φαίνεται να αποδίδουν γενικά καλύτερα για την κατηγοριοποίηση κατά είδος, η οποία είναι και η κατηγοριοποίηση που χρησιμοποιείται πιο συχνά στη βιβλιογραφία. Μια εξαίρεση είναι ο προτεινόμενος αλγόριθμος, ο οποίος αποδίδει συγκριτικά καλά και στην κατηγοριοποίηση κατά συναίσθημα.

LYRIC TEXT MINING IN MUSIC MOOD CLASSIFICATION

Η παρούσα έρευνα εξετάζει τον ρόλο που μπορεί να έχει το λυρικό κείμενο στη βελτίωση της ταξινόμησης της διάθεσης της μουσικής με βάση τον ήχο. Προτείνεται μια νέα μέθοδος για τη δημιουργία ενός μεγάλου συνόλου αληθών δεδομένων, το οποίο περιλαμβάνει 5.585 τραγούδια και 18 κατηγορίες διάθεσης, βασισμένων σε κοινωνικές ετικέτες, με στόχο την αντικατοπτρική παρουσίαση μιας ρεαλιστικής και επικεντρωμένης στον χρήστη προοπτικής. Εξετάστηκε ένα σχετικά πλήρες σύνολο χαρακτηριστικών στίχων και μοντέλων αναπαράστασης. Το καλύτερο σύνολο χαρακτηριστικών στίχων συγκρίθηκε επίσης με ένα κορυφαίο σύστημα βασισμένο στον ήχο. Υπάρχει αυξανόμενο ενδιαφέρον για την ανάπτυξη και αξιολόγηση συστημάτων στο Ανάκτησης Πληροφοριών Μουσικής (Music Information Retrieval) που μπορούν να παρέχουν αυτοματοποιημένη πρόσβαση στη διάσταση της διάθεσης της μουσικής.

ΜΕΘΟΔΟΣ:

ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ

Η αρχή έγινε με μια εσωτερική συλλογή περίπου 21.000 ηχητικών κομματιών. Οι κοινωνικές ετικέτες για αυτά τα τραγούδια συλλέχθηκαν από το last.fm. Τα 12.066 από τα κομμάτια είχαν τουλάχιστον μία ετικέτα από το last.fm. Παράλληλα, στίχοι τραγουδιών συγκεντρώθηκαν από διαδικτυακές βάσεις δεδομένων στίχων. Για την ασφάλεια της ποιότητας των δεδομένων, οι ανιχνευτές χρησιμοποίησαν πληροφορίες τίτλου τραγουδιών, καλλιτέχνη και άλμπουμ για να εντοπίσουν τους σωστούς στίχους. Συνολικά, 8.839 τραγούδια είχαν και ετικέτες και στίχους. Στη συνέχεια, εκτελέστηκε ένα πρόγραμμα αναγνώρισης γλώσσας στους στίχους, και 55

τραγούδια αναγνωρίστηκαν και επιβεβαιώθηκαν χειροκίνητα ως μη-αγγλικά, αφήνοντας στίχους για 8.784 τραγούδια.

ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΚΑΤΗΓΟΡΙΩΝ ΔΙΑΘΕΣΗΣ

Τα δεδομένα από κοινωνικές ετικέτες συχνά είναι θορυβώδη. Για να αφαιρεθούν οι άχρηστες ή χωρίς συναισθηματική σημασία ετικέτες, χρησιμοποιήθηκε το WordNet-Affect, μια επέκταση του WordNet με συναισθηματικές ετικέτες. Από τις 1.586 λέξεις που περιέχει, οι 348 ταυτίστηκαν με τις 61.849 ετικέτες του last.fm. Όμως, δεν συνδέονταν όλες με τη διάθεση στη μουσική. Δύο ειδικοί ερευνητές ανέλαβαν να καθαρίσουν αυτές τις λέξεις, αφαιρώντας όσες δεν σχετίζονταν με συναισθήματα ή ήταν διφορούμενες. Τελικά, παρέμειναν 186 λέξεις, που επισημάνθηκαν σε 4.197 τραγούδια. Οι 186 λέξεις δεν αντιπροσωπεύουν όλες διακριτές έννοιες, καθώς πολλές είναι συνώνυμα και ομαδοποιήθηκαν χρησιμοποιώντας το WordNet-Affect, με αποτέλεσμα να δημιουργηθούν 49 ομάδες ετικετών. Ορισμένες ομάδες ενοποιήθηκαν περαιτέρω λόγω μουσικής ομοιότητας, καταλήγοντας σε 34 κατηγορίες διάθεσης. Για τα πειράματα ταξινόμησης, κρατήθηκαν μόνο οι κατηγορίες με τουλάχιστον 20 τραγούδια, μειώνοντας τις κατηγορίες διάθεσης σε 18 με 135 συνολικά ετικέτες, οι οποίες επικυρώθηκαν από φυσικούς ομιλητές της αγγλικής.

ΕΠΙΛΟΓΗ ΤΩΝ ΤΡΑΓΟΥΔΙΩΝ

Δεν έγινε η επιλογή ενός τραγουδιού για κάποια κατηγορία αν ο τίτλος ή ο καλλιτέχνης του περιλάμβανε τους ίδιους όρους με την κατηγορία. Για παράδειγμα, όλα εκτός από έξι τραγούδια με την ετικέτα "disturbed" άνηκαν στον καλλιτέχνη "Disturbed." Σε αυτήν την περίπτωση, οι χρήστες πιθανώς χρησιμοποίησαν την ετικέτα για να αναφέρουν ξανά τον καλλιτέχνη αντί να περιγράψουν τη διάθεση του τραγουδιού. Για την εξασφάλιση επαρκών δεδομένων για πειράματα βασισμένα στους στίχους, επιλέχθηκαν μόνο τραγούδια των οποίων οι στίχοι περιείχαν πάνω από 100 λέξεις. Μετά την εφαρμογή αυτών των φίλτρων, καταλήξαν με 2.829 μοναδικά τραγούδια. Η ταξινόμηση με πολλαπλές ετικέτες είναι σχετικά νέα στο Ανάκτηση Πληροφοριών Μουσικής (MIR), αλλά όσον αφορά τη διάσταση της διάθεσης, είναι πιο ρεαλιστική από την ταξινόμηση με μία ετικέτα. Ένα μουσικό κομμάτι μπορεί να είναι "χαρούμενο και ήρεμο" ή "επιθετικό και καταθλιπτικό" ταυτόχρονα. Αυτό είναι εμφανές και στο σύνολο δεδομένων, καθώς υπάρχουν πολλά τραγούδια που ανήκουν σε περισσότερες από μία κατηγορίες διάθεσης. Εφαρμόστηκε μια δυαδική προσέγγιση ταξινόμησης για κάθε μία από τις 18 κατηγορίες διάθεσης, και τα 2.829 τραγούδια αποτέλεσαν το σύνολο των θετικών παραδειγμάτων.

ΣΥΜΠΕΡΑΣΜΑ:

Η μελέτη αυτή εξετάζει την αποτελεσματικότητα των χαρακτηριστικών κειμένου στην κατηγοριοποίηση της μουσικής διάθεσης σε 18 διαφορετικές κατηγορίες, οι οποίες προέρχονται από ετικέτες χρηστών. Σε σύγκριση με τις Λειτουργικές Λέξεις και τα Μορφολογικά Στοιχεία, τα Bag-of-Words εξακολουθούν να είναι ο πιο αποδοτικός τύπος χαρακτηριστικού. Η σύγκριση των χαρακτηριστικών των στίχων, του ήχου και των συνδυασμένων χαρακτηριστικών αναδεικνύει μοτίβα που διαφέρουν από τα ευρήματα προηγούμενων ερευνών. Ειδικότερα, τα χαρακτηριστικά των στίχων μπορούν να υπερέχουν των ηχητικών χαρακτηριστικών σε κατηγορίες όπου τα δείγματα είναι πιο περιορισμένα ή όταν οι σημασιολογικές έννοιες των στίχων συνδέονται άμεσα με την κατηγορία διάθεσης. Επιπλέον, ο συνδυασμός στίχων και ήχου βελτιώνει τις επιδόσεις στις περισσότερες, αλλά όχι σε όλες τις κατηγορίες.

A SURVEY ON AUTOMATIC DETECTION OF HATE SPEECH IN TEXT

Η επιστημονική μελέτη του εχθρικού λόγου από την πλευρά της υπολογιστικής επιστήμης είναι σχετικά πρόσφατη. Η παρούσα ανασκόπηση αναλύει και περιγράφει την τρέχουσα κατάσταση του πεδίου, προσφέροντας μια οργανωμένη αναφορά των προηγούμενων προσεγγίσεων, συμπεριλαμβανομένων των κύριων αλγορίθμων, μεθόδων και χαρακτηριστικών που χρησιμοποιούνται. Αυτός ο τομέας έχει αναμφισβήτητο δυναμικό για κοινωνικό αντίκτυπο, ιδίως σε διαδικτυακές κοινότητες και ψηφιακές πλατφόρμες μέσω. Η ανάπτυξη και συστηματοποίηση κοινών πόρων, όπως κατευθυντήριες γραμμές, επισημασμένα σύνολα δεδομένων σε πολλές γλώσσες και αλγόριθμους, είναι ζωτικής σημασίας για την προώθηση της αυτόματης ανίχνευσης του εχθρικού λόγου.

ΜΕΘΟΔΟΣ:

Η μέθοδος είναι δομημένη σε τέσσερις φάσεις:

ΕΠΙΛΟΓΗ ΛΕΞΕΩΝ-ΚΛΕΙΔΙΩΝ

Η πρώτη φάση που πραγματοποιήθηκε ήταν η επιλογή λέξεων-κλειδιών. Εξετάστηκαν όροι που αναφέρονται σε συγκεκριμένους τύπους ρητορικής μίσους (σεξισμός, ρατσισμός και ομοφοβία). Επιπλέον, έλαβαν υπόψη την αναζήτηση για "ρητορική μίσους" και σε άλλες γλώσσες (Πορτογαλικά και Ισπανικά).

ΑΝΑΖΗΤΗΣΗ ΕΓΓΡΑΦΩΝ

Αναζήτησαν σε διάφορες βάσεις δεδομένων και υπηρεσίες (ACM Digital Library, Scopus, Google Scholar και DBLP), με στόχο τη συγκέντρωση του μεγαλύτερου δυνατού αριθμού εγγράφων στους τομείς της πληροφορικής και της μηχανικής. Βάσεις δεδομένων από άλλους επιστημονικούς τομείς δεν λήφθηκαν υπόψη.

ΑΝΑΔΡΟΜΙΚΗ ΑΝΑΖΗΤΗΣΗ

Χρησιμοποιήθηκε το Google Scholar για την εύρεση τόσο για τις αναφορές όσο και για τα έγγραφα που αναφέρουν το αρχικό έργο. Ελέγχθηκαν αυτά τα δύο σύνολα και αναζητήθηκε η έκφραση "ρητορική μίσους" στους τίτλους των υποψήφιων εγγράφων. Αναδρομικά, επαναλήφθηκε η αναζήτηση με τα νέα έγγραφα που βρέθηκαν.

ΦΙΛΤΡΑΡΙΣΜΑ

Πραγματοποιήθηκε ένα αρχικό βήμα φιλτραρίσματος. Έγγραφα από τις κοινωνικές επιστήμες και το δίκαιο εξαιρέθηκαν από τη συλλογή.

ΣΤΡΑΤΗΓΙΚΕΣ:

ΛΕΞΙΚΑ

Μία στρατηγική στην εξόρυξη κειμένου είναι η χρήση λεξικών. Αυτή η προσέγγιση συνίσταται στη δημιουργία μιας λίστας λέξεων που αναζητούνται και μετρώνται στο κείμενο. Αυτές οι συχνότητες μπορούν να χρησιμοποιηθούν άμεσα ως χαρακτηριστικά ή για τον υπολογισμό βαθμολογιών.

BAG-OF-WORDS

Ένα άλλο μοντέλο που μοιάζει με τα λεξικά είναι το bag-of-words. Σε αυτή την προσέγγιση, δημιουργείται ένα σύνολο δεδομένων βασισμένο στις λέξεις που εμφανίζονται στα δεδομένα εκπαίδευσης, αντί να χρησιμοποιείται ένα προκαθορισμένο σύνολο λέξεων, όπως γίνεται με τα λεξικά. Αφού συλλεχθούν όλες οι λέξεις, η συχνότητα της κάθε μιας χρησιμοποιείται ως χαρακτηριστικό για την εκπαίδευση ενός ταξινομητή. Ωστόσο, τα μειονεκτήματα αυτής της μεθόδου περιλαμβάνουν την αγνόηση της σειράς των λέξεων, καθώς και του συντακτικού και σημασιολογικού τους περιεχομένου.

N-GRAMS

Τα N-grams αποτελούν μία από τις πιο ευρέως χρησιμοποιούμενες τεχνικές για την αυτόματη ανίχνευση ρητορικής μίσους και παρόμοιες εφαρμογές. Η συνηθέστερη μέθοδος N-grams περιλαμβάνει τον συνδυασμό διαδοχικών λέξεων σε λίστες μεγέθους N. Επιπλέον, μπορούν να χρησιμοποιηθούν N-grams με χαρακτήρες ή συλλαβές αντί για λέξεις. Αυτή η τεχνική είναι λιγότερο ευαίσθητη στις ορθογραφικές παραλλαγές από ό,τι όταν χρησιμοποιούνται λέξεις. Για την ανίχνευση καταχρηστικής γλώσσας, τα N-grams με χαρακτήρες αποδείχθηκαν πιο αποτελεσματικά από τα N-grams με λέξεις.

TF-IDF

Η μέθοδος TF-IDF (συχνότητα όρου-αντίστροφη συχνότητα εγγράφου) έχει χρησιμοποιηθεί και σε τέτοιου είδους προβλήματα ταξινόμησης. Το TF-IDF μετρά τη σημασία μιας λέξης σε ένα έγγραφο μέσα σε ένα σύνολο εγγράφων και αυξάνεται με βάση τον αριθμό των εμφανίσεων της λέξης στο έγγραφο.

PART-OF-SPEECH

Το POS βοηθά στη βελτίωση της κατανόησης του συμφραζομένου και στην αναγνώριση του ρόλου της λέξης σε μια πρόταση. Αυτές οι μέθοδοι περιλαμβάνουν την κατηγοριοποίηση της λέξης σε τύπους όπως προσωπική αντωνυμία (PRP), ρήμα μη τρίτου ενικού προσώπου του ενεστώτα (VBP), επίθετα (JJ), προσδιοριστικά (DT), και βασικές μορφές ρημάτων (VB). Οι προσεγγίσεις POS έχουν χρησιμοποιηθεί και για την ανίχνευση εχθρικού λόγου. Ωστόσο, το POS έχει αποδειχθεί ότι προκαλεί σύγχυση στην κατηγοριοποίηση όταν χρησιμοποιείται ως χαρακτηριστικό.

ΣΥΜΠΕΡΑΣΜΑ:

Σε αυτή την ανασκόπηση, εξετάστηκε η εξέλιξη της αυτόματης ανίχνευσης ρητορικής μίσους τα τελευταία χρόνια. Έγινε ανάλυση της έννοιας Ρητορικής Μίσους σε διάφορα συμφραζόμενα και προτάθηκε ένας σαφής ορισμός που μπορεί να βοηθήσει στη δημιουργία μοντέλου για την ανίχνευσή της. Διαπιστώθηκε ότι ο προτεινόμενος ορισμός είναι πιο περιεκτικός και περιλαμβάνει και λεπτές μορφές διάκρισης στο διαδίκτυο. Βρέθηκαν τρία διαθέσιμα σύνολα δεδομένων και συγκρίθηκαν οι μελέτες με αλγορίθμους ανίχνευσης ρητορικής μίσους, χωρίς να εντοπιστεί κάποια συγκεκριμένη προσέγγιση που να αποδεικνύεται ανώτερη. Αναφορικά με τα χαρακτηριστικά που χρησιμοποιούνται στις μελέτες, έγινε διαχώριση σε γενικές μεθόδους εξόρυξης κειμένου και ειδικές για την ανίχνευση ρητορικής μίσους. Στις πρώτες περιλαμβάνονται τα N-grams, τα μέρη του λόγου (POS), οι κανόνες και η ανάλυση συναισθήματος. Στις ειδικές μεθόδους για ρητορική μίσους, δόθηκε βάση στη γλώσσα άλλων, την ανωτερότητα της ομάδας και τα στερεότυπα. Παρατηρήθηκε ότι πολλές μελέτες χρησιμοποιούν γενικά χαρακτηριστικά, κάτι που μπορεί να είναι πρόβλημα λόγω της πολυπλοκότητας της ρητορικής μίσους.

ΚΕΦΑΛΑΙΟ Β'

Β' Παραδοσιακές Μέθοδοι

Β'.1 ΠΕΡΙΓΡΑΦΗ

Οι παραδοσιακές μέθοδοι στη μηχανική μάθηση αναφέρονται σε τεχνικές που χρησιμοποιούνταν ευρέως πριν από την ανάπτυξη των βαθιών νευρωνικών δικτύων. Σε αντίθεση με τα μοντέλα βαθιάς μάθησης, οι παραδοσιακές μέθοδοι είναι πιο 'ρηχές', δηλαδή έχουν λιγότερα επίπεδα επεξεργασίας, και συχνά απαιτούν χειροκίνητη εξαγωγή χαρακτηριστικών από τα δεδομένα. Αυτές οι μέθοδοι έχουν αποδειχθεί αποτελεσματικές για πολλές προβληματικές περιοχές και εξακολουθούν να χρησιμοποιούνται σε διάφορες εφαρμογές όπου η βαθιά μάθηση μπορεί να είναι περιττή ή υπερβολικά πολύπλοκη.

Κύριες παραδοσιακές μέθοδοι:

Γραμμική Παλινδρόμηση (Linear Regression)

Χρησιμοποιείται για την πρόβλεψη συνεχόμενων τιμών βάσει μιας γραμμικής σχέσης μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων μεταβλητών. Χρησιμοποιείται κυρίως σε προβλήματα πρόβλεψης με συνεχή δεδομένα.

Λογιστική Παλινδρόμηση (Logistic Regression)

Χρησιμοποιείται για δυαδική ταξινόμηση, δηλαδή για την πρόβλεψη μιας κατηγορίας (ναι/όχι, αληθές/ψευδές). Αντί να προβλέπει συνεχείς τιμές, η λογιστική παλινδρόμηση χρησιμοποιεί τη λογιστική συνάρτηση για να προβλέψει την πιθανότητα μιας εισόδου να ανήκει σε μια συγκεκριμένη κατηγορία.

k-Πλησιέστεροι γείτονες (k-Nearest Neighbors, k-NN)

Είναι μια απλή αλγοριθμική μέθοδος που χρησιμοποιείται για ταξινόμηση και παλινδρόμηση. Το μοντέλο αναζητά τα k πιο κοντινά δείγματα δεδομένων σε ένα σημείο και κατηγοριοποιεί το νέο δεδομένο βασιζόμενο σε αυτά.

Δέντρα Αποφάσεων (Decision Trees)

Αυτή η μέθοδος δημιουργεί ένα δενδροειδές μοντέλο που χωρίζει τα δεδομένα σε κόμβους και κλαδιά με βάση χαρακτηριστικά και κανόνες απόφασης. Κάθε φύλλο του δέντρου αντιπροσωπεύει μια κατηγορία ή τιμή που πρέπει να προβλεφθεί.

Τυχαία Δάση (Random Forests)

Είναι ένας αλγόριθμος που συνδυάζει πολλά δέντρα απόφασης (ensemble learning). Η τελική πρόβλεψη βασίζεται στη 'ψηφοφορία' από πολλά δέντρα απόφασης. Αυτός ο αλγόριθμος μειώνει τον κίνδυνο υπερεκπαίδευσης (overfitting) που μπορεί να εμφανιστεί στα απλά δέντρα αποφάσεων.

Διανύσματα Υποστήριξης (Support Vector Machines, SVMs)

Είναι μια μέθοδος για ταξινόμηση, παλινδρόμηση και ανάλυση ανωμαλιών. Χρησιμοποιεί υπερ-επίπεδα για να διαχωρίσει τα δεδομένα σε διαφορετικές κατηγορίες. Σκοπός είναι η μεγιστοποίηση της απόστασης μεταξύ των κοντινότερων σημείων δεδομένων από διαφορετικές κατηγορίες.

Naive Bayes

Βασίζεται στο θεώρημα του Bayes και χρησιμοποιείται συχνά για προβλήματα ταξινόμησης. Ο Bayes θεωρεί ότι τα χαρακτηριστικά είναι ανεξάρτητα, κάτι που δεν είναι πάντα ρεαλιστικό αλλά συχνά παρέχει αποτελεσματικά αποτελέσματα.

Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis, PCA)

Είναι μια τεχνική για μείωση της διάστασης (dimensionality reduction). Εντοπίζει τις κύριες διαστάσεις των δεδομένων και τα απεικονίζει σε έναν χώρο χαμηλότερης διάστασης, διατηρώντας τη μεγαλύτερη δυνατή πληροφορία.

k-Μέσοι Όροι (k-Means Clustering)

Είναι μια μέθοδος μη επιβλεπόμενης μάθησης για την ομαδοποίηση δεδομένων σε k ομάδες, όπου κάθε σημείο δεδομένων ανήκει στο κέντρο της ομάδας με το οποίο έχει την μικρότερη απόσταση.

Χαρακτηριστικά των παραδοσιακών μεθόδων:

Χειροκίνητη Εξαγωγή Χαρακτηριστικών

Οι παραδοσιακές μέθοδοι συνήθως απαιτούν χειροκίνητη παρέμβαση για την εξαγωγή χαρακτηριστικών που είναι χρήσιμα για την εκπαίδευση των μοντέλων.

Χαμηλή Υπολογιστική Ισχύς

Σε σχέση με τα βαθιά νευρωνικά δίκτυα, οι παραδοσιακές μέθοδοι είναι πιο "ελαφριές" και απαιτούν λιγότερη υπολογιστική ισχύ και χρόνο για να εκπαιδευτούν.

Χρήση σε Μικρότερα Δεδομένα

Είναι συνήθως πιο αποτελεσματικές σε μικρότερα datasets, όπου τα βαθιά νευρωνικά δίκτυα μπορεί να είναι υπερβολικά και να μην αποδίδουν καλά.

Παρότι οι βαθιές νευρωνικές δομές έχουν γίνει εξαιρετικά δημοφιλείς τα τελευταία χρόνια, οι παραδοσιακές μέθοδοι εξακολουθούν να είναι χρήσιμες, ειδικά σε περιπτώσεις όπου τα δεδομένα είναι περιορισμένα ή η πολυπλοκότητα του προβλήματος δεν απαιτεί βαθιά μάθηση.

B'.2 ΕΡΕΥΝΑ ΚΛΑΣΙΚΩΝ ΜΕΘΟΔΩΝ

DETECTION OF OBJECTIONABLE SONG LYRICS USING WEAKLY SUPERVISED LEARNING AND NATURAL LANGUAGE PROCESSING TECHNIQUES

Η εξάπλωση των ψηφιακών και κοινωνικών μέσων έχει συμβάλλει σημαντικά στην αυξημένη διαθεσιμότητα της μουσικής και των στίχων της. Ωστόσο, έχει παρατηρηθεί ότι το προσβλητικό περιεχόμενο αποτελεί σοβαρό ζήτημα στη σύγχρονη ψηφιακή εποχή. Η συγκεκριμένη έρευνα επικεντρώνεται στην ανάπτυξη ενός συστήματος για τον αυτόματο εντοπισμό στίχων που περιλαμβάνουν προσβλητικά ή αμφιλεγόμενα θέματα, όπως η ρητορική μίσους, ο σεξισμός, ο ρατσισμός κ.α. Για τον σκοπό αυτό, χρησιμοποιήθηκαν ευριστικοί κανόνες με το πλαίσιο Snorkel για την παραγωγή ετικετών, ενώ αξιολογήθηκε η απόδοση διάφορων αλγορίθμων μηχανικής μάθησης, όπως η Λογιστική Παλινδρόμηση, η Random Forest και η Μηχανή Υποστήριξης Διανυσμάτων (SVM), στα δεδομένα αυτά.

ΜΕΘΟΔΟΣ:

ΣΥΛΛΟΓΗ ΚΑΙ ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Τα κύρια σύνολα δεδομένων που χρησιμοποιήθηκαν σε αυτή τη μελέτη εισήχθησαν από δύο ξεχωριστά αρχεία: lyrics-data.csv και artists-data.csv. Αυτά τα αρχεία περιλαμβάνουν δεδομένα που αφορούν τους στίχους των τραγουδιών και τους καλλιτέχνες αντίστοιχα. Έπειτα, γίνεται η αφαίρεση αχρείαστων συμβόλων, χαρακτήρων ή θορύβου από τα δεδομένα, όπως σημεία στίξης, κενά ή μη χρήσιμες λέξεις. Τέλος, η διαδικασία κατακερματισμού των στίχων σε μικρότερα μέρη, όπως λέξεις ή φράσεις, για ανάλυση.

COUNT VECTORIZER

Η μέθοδος Count Vectorizer είναι η πιο απλή στην έρευνα αυτή. Αναπαριστά κάθε δεδομένο με έναν πίνακα one-hot που δείχνει αν μια λέξη ή συνδυασμός λέξεων υπάρχει στο κείμενο του τραγουδιού. Χρησιμοποιήθηκε η υλοποίηση CountVectorizer του Scikit-Learn, με περιοχή n-gram (1,3) για καλύτερα αποτελέσματα. Αν και είναι απλή στην υλοποίηση, δημιουργεί μεγάλο και αραιό πίνακα χαρακτηριστικών.

TF-IDF

Για τη δημιουργία ενσωματώσεων TF-IDF χρησιμοποιήθηκε η κλάση TfidfVectorizer του Scikit-Learn, που επιλέχθηκε μέσω βελτιστοποίησης. Εστίασε σε λέξεις με περισσότερους από έναν χαρακτήρες και περιορίστηκε σε 10.000 χαρακτηριστικά. Αν και δημιουργεί πιο πυκνό πίνακα χαρακτηριστικών σε σύγκριση με το CountVectorizer, δεν εξετάζει το συμφραζόμενο της λέξης.

DistilBERT

Οι ενσωματώσεις DistilBERT προέρχονται από το μοντέλο DistilBERT, μια ελαφρύτερη έκδοση του BERT. Εξασφαλίζουν αποτελεσματική αποτύπωση συμφραζόμενων λέξεων και βελτιώνουν διάφορες γλωσσικές εργασίες. Το DistilBERT είναι πιο αποδοτικό λόγω της μειωμένης του πολυπλοκότητας σε σύγκριση με το BERT.

ΜΟΝΤΕΛΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Για την πρόβλεψη ετικετών κλάσεων, χρησιμοποιήθηκαν τρεις κλασικοί ταξινομητές μηχανικής μάθησης: Support Vector Machine, Random Forest και Logistic Regression, με τις υλοποιήσεις τους από το Scikit-Learn. Η Logistic Regression χρησιμοποιεί το liblinear με όριο 1000 επαναλήψεων. Για το Random Forest, ορίστηκε η τυχαία κατάσταση στο 1 και ο αριθμός των δέντρων σε 50. Η ανάλυση Support Vector Machine έγινε με την υλοποίηση LinearSVC και τυχαία κατάσταση στο μηδέν.

ΣΥΜΠΕΡΑΣΜΑ:

Τα αποτελέσματα έδειξαν ότι το DistilBert υπερτερεί έναντι των παραδοσιακών μεθόδων που βασίζονται στη συχνότητα (TF-IDF και CountVectorizer) κατά 3-5% στη βαθμολογία F1. Ακόμη πιο εντυπωσιακά ήταν τα αποτελέσματα σε σύνολα δεδομένων, όπου η Λογιστική Παλινδρόμηση και η SVM με δεδομένα DistilBert εμφάνισαν παρόμοια απόδοση, επιτυγχάνοντας αύξηση 6-8% στη βαθμολογία F1 σε σύγκριση με τις μεθόδους που βασίζονται στη συχνότητα. Τέλος, η μελέτη δείχνει ότι η προτεινόμενη μέθοδος μπορεί να χρησιμοποιηθεί για την ανίχνευση προσβλητικών στίχων και σε άλλες γλώσσες, εφόσον χρησιμοποιηθούν παρόμοια δεδομένα. Η εργασία μας μπορεί να αποτελέσει χρήσιμο εργαλείο για τους επαγγελματίες της μουσικής βιομηχανίας και τα συστήματα παρακολούθησης ψηφιακού περιεχομένου στην προσπάθειά τους να ανιχνεύσουν ρητορική μίσους και προσβλητικό περιεχόμενο, υιοθετώντας τις μεθοδολογίες και τις προσεγγίσεις που αναπτύχθηκαν.

AUDIO-BASED DETECTION OF EXPLICIT CONTENT IN MUSIC

Σε αυτή την εργασία, προτείνεται ένα καινοτόμο αυτόματο σύστημα για την ανίχνευση ρητού περιεχομένου απευθείας στο ηχητικό σήμα. Η μεθοδολογία αποτελείται από μια αρθρωτή προσέγγιση, που περιλαμβάνει ένα μοντέλο μετατροπής ήχου σε χαρακτήρες, ένα σύστημα εντοπισμού λέξεων-κλειδιών που βασίζεται σε ένα προσεκτικά κατασκευασμένο λεξικό λέξεων-κλειδιών, και ένα μοντέλο Random Forest για την τελική απόφαση. Κατά την γνώση μας, πρόκειται για το πρώτο σύστημα που ανιχνεύει ρητό περιεχόμενο χρησιμοποιώντας αποκλειστικά ηχητικά δεδομένα. Εμφανίζεται η απόδοση κάθε επιμέρους μονάδας σε συγκεκριμένα υπο-καθήκοντα και γίνεται σύγκριση της προσέγγισής αυτής με ένα σύστημα βασισμένο στους στίχους καθώς και με μια πιο απλή αρχιτεκτονική, από άκρη σε άκρη. Τα αποτελέσματα είναι ελπιδοφόρα, με F1-score 67% σε ένα σύνολο δεδομένων βιομηχανικής κλίμακας για την ανίχνευση ρητού περιεχομένου.

ΜΕΘΟΔΟΣ:

ΑΝΑΓΝΩΡΙΣΗ AUDIO-TO-CHARACTER

Το μοντέλο της παρούσας έρευνας, στο οποίο στηρίζεται η μονάδα A2C (Audio-to-Character), έχει υλοποιηθεί με ένα Συγκλητικό Επαναλαμβανόμενο Νευρωνικό Δίκτυο (CRNN), το οποίο εκπαιδεύεται χρησιμοποιώντας τον αλγόριθμο Connectionist Temporal Classification (CTC). Το CTC έχει ήδη χρησιμοποιηθεί με επιτυχία στην αναγνώριση ομιλίας. Για να μειωθεί η πολυπλοκότητα των χαρακτηριστικών και να επιταχύνει η διαδικασία εκπαίδευσης, γίνεται πρόσθεση επιπλέον συγκλητικών επιπέδων. Οι επαναλαμβανόμενες στρώσεις που επιλέχθηκαν είναι αμφίδρομα LSTM, επιτρέποντας στις εξόδους να εξαρτώνται από ολόκληρη την ακολουθία εισόδου σε κάθε χρονικό πλαίσιο.

Ο αλγόριθμος CTC δίνει τη δυνατότητα εκπαίδευσης RNN μοντέλων χωρίς να απαιτείται απόλυτη ευθυγράμμιση των δεδομένων. Εισάγεται ένας ειδικός χαρακτήρας “κενό” που αντιπροσωπεύει την απουσία εκπομπής. Το μοντέλο μπορεί να εκπέμπει οποιονδήποτε χαρακτήρα, συμπεριλαμβανομένου του “κενού”, σε κάθε χρονικό πλαίσιο. Η συνολική πιθανότητα της παραγόμενης ακολουθίας χαρακτήρων μεγιστοποιείται με τη χρήση του CTC αλγορίθμου, υπολογίζοντας όλες τις πιθανές ευθυγραμμίσεις για την είσοδο.

ΑΝΑΓΝΩΡΙΣΗ ΛΕΞΕΩΝ-ΚΛΕΙΔΙΩΝ

Η αναγνώριση λέξεων-κλειδιών γίνεται μέσω μιας συνάρτησης αποκωδικοποίησης βασισμένης στην CTC. Η λέξη-κλειδί τροποποιείται προσθέτοντας ένα “κενό” σύμβολο στην αρχή, στο τέλος και ανάμεσα στους χαρακτήρες. Το σύστημα προσπαθεί να βρει το μονοπάτι που μεγιστοποιεί το σκορ CTC για αυτή τη λέξη. Οι υπολογισμοί γίνονται αποδοτικά με έναν αλγόριθμο εμπρός-πίσω, και επιτρέπονται μεταβάσεις μόνο μεταξύ κενών και μη κενών χαρακτήρων, καθώς και μεταξύ διαφορετικών μη κενών χαρακτήρων.

ΜΟΝΤΕΛΟ ΑΝΑΓΝΩΡΙΣΗΣ:

Το κύριο συστατικό του μοντέλου A2C εκπαιδεύεται χρησιμοποιώντας το σύνολο δεδομένων DALI, το οποίο περιέχει 5.358 ηχητικά κομμάτια με ευθυγραμμισμένους στίχους σε επίπεδο παραγράφου, γραμμής και λέξης, καλύπτοντας διάφορα είδη δυτικής μουσικής όπως ροκ, ραπ και ηλεκτρονική.

Τα ηχητικά κομμάτια τμηματοποιούνται σε παράθυρα διάρκειας 5 δευτερολέπτων, με βήμα 2,5 δευτερόλεπτα, δημιουργώντας εκπαιδευτικά δείγματα. Οι ακολουθίες χαρακτήρων συντίθενται από τις λέξεις που ξεκινούν εντός του τμήματος και, αν δεν υπάρχουν λέξεις, εισάγεται ο ειδικός χαρακτήρας "I" για να δηλώσει "Instrumental".

Το μοντέλο περιλαμβάνει δύο συγκλιτικά επίπεδα ακολουθούμενα από τρία επίπεδα LSTM δύο κατευθύνσεων, καθώς και ένα πυκνό επίπεδο στο τέλος. Οι τιμές που εξάγονται περιλαμβάνουν συντελεστές φίλτρων λογαριθμικής κλίμακας και ενέργεια, καθώς και τις παράγωγες πρώτης και δεύτερης τάξης.

Η εκπαίδευση πραγματοποιείται με χρήση της απώλειας CTC, εφαρμόζοντας τον αλγόριθμο ADAM για την ελαχιστοποίηση της απώλειας. Το μοντέλο επιτυγχάνει ποσοστό σφάλματος χαρακτήρων (CER) 47,41%, το οποίο είναι συγκρίσιμο με τα αποτελέσματα άλλων μελετών σε διαφορετικά σύνολα δεδομένων.

ΑΝΙΧΝΕΥΣΗ ΛΕΞΕΩΝ-ΚΛΕΙΔΙΩΝ:

Για την παραγωγή του λεξικού, υπολογίζεται η σημασία, η οποία βασίζεται στη συχνότητα της λέξης σε ρητούς και μη ρητούς στίχους. Απορρίπτονται λέξεις-κλειδιά και άλλες λιγότερο σημαντικές. Το λεξικό περιλαμβάνει 128 λέξεις με τη μεγαλύτερη σημασία.

Η απόδοση της αποκωδικοποίησης στο σύνολο δοκιμής DALI δείχνει ότι το 75% των λέξεων-κλειδιών έχουν AUC μεγαλύτερο από 0.81. Δεδομένου ότι είναι η πρώτη φορά που υπολογίζονται τέτοιες μετρήσεις για μουσική, δεν είναι δυνατή η σύγκριση με άλλα αποτελέσματα. Ο διανυσματικός χαρακτήρας περιέχει πληροφορίες για την παρουσία λέξεων-κλειδιών.

Η θέση του χαρακτήρα "κενό" στη γραμμή αποκωδικοποίησης είναι πολύ κοντά στην αληθινή θέση της λέξης, υποδεικνύοντας ότι το μοντέλο εντοπίζει σωστά τις θέσεις των λέξεων, κάτι που συμφωνεί με τα αποτελέσματα ευθυγράμμισης στίχων με ήχο.

ΑΝΙΧΝΕΥΣΗ ΡΗΤΩΝ ΣΤΙΧΩΝ:

Χρησιμοποιείται ένα μοντέλο που έχει εκπαιδευτεί σε ένα ιδιωτικό σύνολο δεδομένων που είναι επισημασμένο ως ρητό ή μη ρητό. Για να μειωθεί η υπερπροσαρμογή, αποκλείονται κομμάτια από το σύνολο δεδομένων DALI. Η κατανομή των μουσικών ειδών είναι σημαντικά διαφορετική μεταξύ ρητών και μη ρητών κομματιών, με το ραπ να είναι υπέρ-παρουσιασμένο στο ρητό περιεχόμενο. Για να αντιμετωπιστεί αυτό, το σύνολο δεδομένων ισορροπείται κατά είδος, με αποτέλεσμα 2.600 μη ρητά και 2.530 ρητά κομμάτια.

Μια κατανομή που λαμβάνει υπόψη τους καλλιτέχνες χωρίζει τα δεδομένα σε σύνολα εκπαίδευσης (70%), επικύρωσης (15%) και δοκιμής (15%). Ένα δευτερεύον σύνολο δεδομένων για τη δημιουργία λεξικού περιλαμβάνει 24.250 κομμάτια και για τις δύο κατηγορίες (ρητά και μη ρητά), διασφαλίζοντας ότι δεν υπάρχει επικάλυψη.

Το μοντέλο συγκρίνεται με δύο βάσεις αναφοράς: ένα Συνελικτικό Αναδρομικό Νευρωνικό Δίκτυο (CRNN) που ταξινομεί τον ήχο άμεσα και μια μέθοδο που βασίζεται σε λεξικό, η οποία ταξινομεί τα τραγούδια με βάση ρητές λέξεις-κλειδιά στους στίχους. Το CRNN χρησιμοποιεί φίλτρα mel-scale log και εκπαιδεύεται για 3.000 εποχές, ενώ η μέθοδος λεξικού βασίζεται σε προεπιλεγμένες ρητές λέξεις-κλειδιά.

Για το μοντέλο ανίχνευσης, χρησιμοποιείται ένας ταξινομητής Random Forest, με υπερπαραμέτρους που βελτιστοποιούνται μέσω τυχαίας αναζήτησης. Η αξιολόγηση επικεντρώνεται στην ακρίβεια, την ανάκληση και το F1-score, με ιδιαίτερη έμφαση στην υψηλή ανάκληση λόγω της ευαισθησίας του ρητού περιεχομένου.

ΣΥΜΠΕΡΑΣΜΑ:

Η έρευνα αυτή ασχολείται με την καινοτόμο πρόκληση της ανίχνευσης ρητού μουσικού περιεχομένου αποκλειστικά μέσω ηχητικών δεδομένων. Αυτή η διαδικασία είναι ιδιαίτερα απαιτητική, καθώς η αναγνώριση ρητών στίχων από τον ήχο χωρίς τη βοήθεια των στίχων ή άλλων ενδείξεων είναι δύσκολη. Ωστόσο, η προτεινόμενη αρθρωτή προσέγγιση που εφαρμόζουμε αποδίδει υποσχόμενα αποτελέσματα, δείχνοντας ότι η μέθοδος είναι ικανή να αναγνωρίσει ρητό περιεχόμενο με ικανοποιητική ακρίβεια.

Ένα σημαντικό χαρακτηριστικό του συστήματος είναι ότι οι αποφάσεις που λαμβάνει μπορούν να εξηγηθούν μέσω της πιθανότητας παρουσίας συγκεκριμένων λέξεων-κλειδιών. Αυτό είναι κρίσιμο, δεδομένης της ευαισθησίας του αντικειμένου, καθώς επιτρέπει στους χρήστες και τους ερευνητές να κατανοήσουν τους λόγους πίσω από τις αποφάσεις του συστήματος και να έχουν εμπιστοσύνη στα αποτελέσματα.

Για το μέλλον, μπορεί να γίνει μεγαλύτερη εξερεύνηση της ενίσχυσης της διαδικασίας αποκωδικοποίησης λέξεων-κλειδιών, χρησιμοποιώντας γλωσσικά μοντέλα σε επίπεδο χαρακτήρα. Αυτή η προσέγγιση θα μπορούσε να βελτιώσει περαιτέρω την ικανότητα του συστήματος να αναγνωρίζει ρητό περιεχόμενο, προσφέροντας μια πιο ακριβή και αξιόπιστη ανίχνευση.

HATE SPEECH CLASSIFICATION USING SVM AND NAIVE BAYES

Η διάδοση του μίσους, που στο παρελθόν περιοριζόταν στις προφορικές επικοινωνίες, έχει μεταφερθεί ταχύτατα στον διαδικτυακό χώρο, στη μουσική κ.α. Πολλές χώρες έχουν θεσπίσει νομοθεσίες για την καταπολέμηση της ρητορικής μίσους. Ωστόσο, όσο αυξάνεται το περιεχόμενο στο διαδίκτυο (σε φορουμ, τραγούδια σε διαδικτυακές πλατφόρμες κ.α) αυξάνεται και η διάδοση της ρητορικής μίσους. Η χειροκίνητη αξιολόγηση του μίσους σε διαδικτυακές πλατφόρμες είναι ανέφικτη λόγω του τεράστιου όγκου δεδομένων, καθώς απαιτεί πόρους και χρόνο. Κατά συνέπεια, κρίνεται απαραίτητη η αυτόματη επεξεργασία του περιεχομένου για τον εντοπισμό και την απομάκρυνση της ρητορικής μίσους. Πολλές πρόσφατες προσεγγίσεις αντιμετωπίζουν προβλήματα διαφάνειας, καθιστώντας δύσκολη την κατανόηση των αποφάσεων που λαμβάνονται από τα συστήματα. Σε αυτή τη μελέτη, προτείνονται λύσεις για τον αυτόματο εντοπισμό εχθρικού λόγου με τη χρήση των αλγορίθμων SVM και Naïve Bayes. Οι αλγόριθμοι αυτοί επιτυγχάνουν απόδοση κοντά στην κορυφαία τεχνολογία, ενώ παραμένουν απλούστεροι και προσφέρουν πιο κατανοητές αποφάσεις σε σύγκριση με άλλες μεθόδους. Η αξιολόγηση έδειξε ακρίβεια ταξινόμησης περίπου 99% για τον SVM και 50% για τον Naïve Bayes στο σύνολο δοκιμών.

ΜΕΘΟΔΟΣ:

Η μέθοδος CRISP-DM επιλέχθηκε για αυτή την έρευνα, καθώς περιλαμβάνει όλα τα απαραίτητα βήματα για την επίτευξη των στόχων της. Αποτελεί μια ολοκληρωμένη μεθοδολογία εξόρυξης δεδομένων και παρέχει ένα πλήρες πλαίσιο για την υλοποίηση ενός έργου στον τομέα αυτό. Για την ανίχνευση και κατηγοριοποίηση της ρητορικής μίσους, χρησιμοποιήθηκαν αλγόριθμοι επιτηρούμενης μάθησης, όπως οι SVM και ο Naïve Bayes, με τη βοήθεια των εργαλείων μηχανικής μάθησης WEKA, της γλώσσας προγραμματισμού Java και του IDE NetBeans.

Ακολουθήθηκαν τα εξής βήματα για την επίτευξη του επιθυμητού αποτελέσματος:

- i) Συλλογή δείγματος δεδομένων με βάση την κατηγοριοποίηση μάθησης.
- ii) Προεπεξεργασία (δηλαδή, το σύνολο δεδομένων εφοδιάστηκε με ετικέτες), καθώς πρόκειται για προσέγγιση επιτηρούμενης μάθησης.
- iii) Εφαρμογή εξαγωγής χαρακτηριστικών με τη βιβλιοθήκη Weka (για μετατροπή των χαρακτηριστικών σε αριθμητικές τιμές για ανάλυση κατηγοριοποίησης).
- iv) Επαναδειγματοληψία του συνόλου δεδομένων εφαρμόζοντας σύνολο εκπαίδευσης και σύνολο δοκιμών κατά την ανάλυση ανάπτυξης συστήματος με εργαλεία Weka.
- v) Καθαρισμός δεδομένων δηλαδή αφαιρούνται θορυβώδη δεδομένα και επαναλαμβανόμενες εγγραφές. Η διαδικασία είναι χρονοβόρα και επιρρεπής σε σφάλματα, οπότε αναπτύχθηκε εργαλείο Java για αυτοματοποίηση και μείωση του μεγέθους των δεδομένων με το Weka.
- vi) Μετασχηματισμός δεδομένων δηλαδή χρειάζονται τεχνικές όπως η κανονικοποίηση και η μετατροπή από ονοματικές σε αριθμητικές τιμές για καλύτερη απόδοση αλγορίθμων, όπως ο Naïve Bayes.
- vii) Μείωση δεδομένων δηλαδή ότι επικεντρώνεται στην επιλογή σχετικών χαρακτηριστικών και στην αφαίρεση άσχετων, βελτιώνοντας την πρόβλεψη και μειώνοντας τον χρόνο εκπαίδευσης.

ΑΛΓΟΡΙΘΜΟΙ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΓΙΑ ΤΗΝ ΤΑΞΙΝΟΜΗΣΗ ΚΑΙ ΤΟΝ ΕΝΤΟΠΙΣΜΟ ΤΗΣ ΡΗΤΟΡΙΚΗΣ ΜΙΣΟΥΣ:

Support Vector Machine

Ο αλγόριθμος SVM είναι ένας από τους επιτηρούμενους αλγόριθμους μηχανικής μάθησης που χρησιμοποιείται για διάφορα προβλήματα κατηγοριοποίησης. Έχει εφαρμογές στην ανάλυση πιστωτικού κινδύνου, στη ιατρική διάγνωση, στην κατηγοριοποίηση κειμένου και στην εξαγωγή πληροφοριών. Ο SVM είναι ιδιαίτερα κατάλληλος για δεδομένα υψηλής διάστασης. Υπάρχουν πολλοί λόγοι που υποστηρίζουν αυτή την άποψη. Συγκεκριμένα, η πολυπλοκότητα των ταξινομητών εξαρτάται από τον αριθμό των υποστηρικτικών διανυσμάτων αντί για τις διαστάσεις των δεδομένων, παράγουν το ίδιο υπερ-επίπεδο για επαναλαμβανόμενα σύνολα εκπαίδευσης και έχουν καλύτερες ικανότητες γενίκευσης.

Naïve Bayes

Αυτοί είναι πιθανοί ταξινομητές που χρησιμοποιούνται συχνά στη μηχανική μάθηση. Αυτοί οι ταξινομητές, αν και στατιστικοί, έχουν επίσης τη δυνατότητα μάθησης. Ο Naïve Bayes χρησιμοποιεί ένα πολυωνυμικό μοντέλο για μεγάλα σύνολα δεδομένων, και η απόδοση μπορεί να βελτιωθεί με την αναζήτηση εξαρτήσεων μεταξύ χαρακτηριστικών. Χρησιμοποιείται κυρίως στην προεπεξεργασία δεδομένων λόγω της ευκολίας στους υπολογισμούς. Η λογική του και η πιθανοτική συμπερασματολογία συμβάλλουν στην πρόβλεψη της στοχευμένης κατηγορίας, με τα χαρακτηριστικά να διαδραματίζουν κρίσιμο ρόλο στην κατηγοριοποίηση. Έτσι, η αποτελεσματικότητα του Naïve Bayes εξαρτάται από την ακρίβεια των εκτιμήσεων των όρων πιθανότητας, κάτι που είναι δύσκολο όταν τα

δεδομένα εκπαίδευσης είναι περιορισμένα. Η μέθοδος Naïve Bayes βασίζεται στο θεώρημα του Bayes και είναι πιο αποτελεσματική σε δεδομένα υψηλής διάστασης. Ο 'βαγιανικός' ταξινομητής υπολογίζει την πιο πιθανή έξοδο με βάση την είσοδο, και η προσθήκη νέων δεδομένων σε πραγματικό χρόνο δεν αποτελεί πρόβλημα, επιτρέποντας τη δημιουργία ενός πιο ακριβούς πιθανοτικού ταξινομητή.

ΣΥΜΠΕΡΑΣΜΑ:

Πρόσφατες μελέτες αποκαλύπτουν ότι έχουν προταθεί πολλές τεχνικές για την ανίχνευση ρητορικής μίσους. Οι περισσότερες έρευνες σε αυτόν τον τομέα περιορίζονται σε συγκεκριμένα χαρακτηριστικά, γεγονός που υποδεικνύει την ανάγκη για την αναζήτηση νέων παραμέτρων. Η συμπεριφορά της ρητορικής μίσους εξετάστηκε, με σκοπό την ανάπτυξη ενός συστήματος ανίχνευσης και κατηγοριοποίησης χρησιμοποιώντας τους αλγόριθμους Naïve Bayes και SVM. Το πρόγραμμα αναπτύχθηκε με τη γλώσσα Java και χρησιμοποιήθηκαν εργαλεία Weka για ταξινόμηση. Τα δεδομένα συλλέχθηκαν από πηγές που ονομάζονται Unique Client Identifier, οι οποίες είναι χρήσιμες για την εμπειρική ανάλυση αλγορίθμων μηχανικής μάθησης. Επιπλέον, οι υπερπαραμέτροι των SVM και Naïve Bayes ρυθμίστηκαν ώστε να επιτευχθούν οι καλύτερες επιδόσεις. Διαπιστώθηκε ότι ο SVM παρουσίασε καλύτερα αποτελέσματα σε σύγκριση με τον Naïve Bayes κατά τη διάρκεια των πειραμάτων. Η ανίχνευση χωρίστηκε σε δύο κατηγορίες: προσβλητικές και μη προσβλητικές ετικέτες. Στο μέλλον, οι ερευνητές θα πρέπει να συγκρίνουν και άλλους αλγορίθμους, όπως το Random Forest, τον αλγόριθμο k-NN, τη βαθιά μάθηση και τα Τεχνητά Νευρωνικά Δίκτυα, προκειμένου να βρουν τον πιο αποτελεσματικό για την ανίχνευση ρητορικής μίσους.

SOME EFFECTIVE TECHNIQUES FOR NAIVE BAYES TEXT CLASSIFICATION

Ο αλγόριθμος Naive Bayes είναι αποτελεσματικός σε πολλές εργασίες εξόρυξης δεδομένων, αλλά εμφανίζει απογοητευτικά αποτελέσματα στο ζήτημα της αυτόματης κατηγοριοποίησης κειμένου. Από την παρατήρηση της απόδοσής του σε κείμενα φυσικής γλώσσας, ανακαλύψαμε ένα σημαντικό πρόβλημα στη διαδικασία εκτίμησης παραμέτρων, το οποίο οδηγεί σε χαμηλή απόδοση στην κατηγοριοποίηση κειμένων. Σε αυτή την εργασία, προτείνονται δύο εμπειρικές ευρετικές μέθοδοι: κανονικοποίηση κειμένου ανά έγγραφο και ζύγιση χαρακτηριστικών. Αν και αυτές οι μέθοδοι είναι σχετικά ad hoc, ο προτεινόμενος ταξινομητής κειμένου Naive Bayes αποδίδει εξαιρετικά καλά στις τυπικές συλλογές αναφοράς, ανταγωνιζόμενος σύγχρονους ταξινομητές κειμένου που χρησιμοποιούν πολύπλοκες μεθόδους μάθησης, όπως η SVM.

ΜΕΘΟΔΟΣ:

ΠΟΛΥΔΙΑΣΤΑΤΟ ΜΟΝΤΕΛΟ POISSON ΓΙΑ ΤΑΞΙΝΟΜΗΣΗ ΚΕΙΜΕΝΟΥ

Η κατανομή Poisson χρησιμοποιείται ευρέως για την μοντελοποίηση του αριθμού εμφανίσεων ενός φαινομένου σε μια συγκεκριμένη χρονική ή χωρική περίοδο, όπως ο αριθμός τηλεφωνικών κλήσεων σε ένα κέντρο εξυπηρέτησης ή ο αριθμός ελαττωμάτων σε μια συγκεκριμένη ποσότητα υφάσματος ή χαρτιού. Το μοντέλο Poisson έχει εξεταστεί εκτενώς στη βιβλιογραφία της Πληροφορίας, δεδομένου ότι η συχνότητα εμφάνισης μιας λέξης σε ένα έγγραφο μπορεί να περιγραφεί με αυτή την κατανομή. Ωστόσο, σπάνια έχει μελετηθεί στο πλαίσιο της κατηγοριοποίησης κειμένου, γεγονός που μας παρακινεί να διερευνήσουμε τη χρήση του, για την εκπαίδευση του ταξινομητή Naive Bayes.

ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ

Από τον ορισμό της κατανομής Poisson, η παράμετρος Poisson υποδεικνύει τον μέσο αριθμό γεγονότων κατά την παρατήρηση μιας σταθερής περιόδου. Επομένως, οι παράμετροι Poisson του μοντέλου κατηγοριοποίησης κειμένου μας υποδεικνύουν πόσες φορές εμφανίζεται ένας όρος στα θετικά και αρνητικά έγγραφα εκπαίδευσης κατά μέσο όρο.

ΕΠΙΛΟΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Η επιλογή χαρακτηριστικών είναι ένα κρίσιμο ζήτημα στην κατηγοριοποίηση κειμένου, διότι μπορεί να μειώσει την πολυπλοκότητα των αλγορίθμων μηχανικής μάθησης. Οι Yang και Pedersen μελέτησαν διάφορες μεθόδους για την επιλογή χρήσιμων χαρακτηριστικών, όπως η αμοιβαία πληροφορία, η κερδοφορία πληροφορίας και οι στατιστικές κ.λπ. Είναι γεγονός ότι η χρήση όλων των χαρακτηριστικών λέξεων συνήθως ενισχύει την απόδοση του ταξινομητή. Ο Joachims υποστήριξε ότι δεν υπάρχουν άχρηστα χαρακτηριστικά όρων και ότι είναι προτιμότερο να αξιοποιούνται όλα τα χαρακτηριστικά λέξεων. Στην περίπτωση της πολυδιάστατης κατηγοριοποίησης κειμένου με τον αλγόριθμο Naive Bayes, έχει αποδειχθεί ότι η επιλογή χαρακτηριστικών μπορεί να βελτιώσει την απόδοση κατηγοριοποίησης σε ορισμένες κατηγορίες. Ωστόσο, αυτή η βελτίωση προκύπτει από τη συσσώρευση σφαλμάτων λόγω της κακής εκτίμησης παραμέτρων του πολυδιάστατου μοντέλου Naive Bayes.

ΣΥΜΠΕΡΑΣΜΑ:

Σε αυτή την έρευνα, προτείνεται ένα μοντέλο κατηγοριοποίησης κειμένου Naive Bayes που βασίζεται σε κατανομή Poisson και περιλαμβάνει μέθοδο ενίσχυσης-βαρών. Το μοντέλο αυτό υποθέτει ότι κάθε έγγραφο δημιουργείται από ένα πολυδιάστατο μοντέλο Poisson. Επίσης, προτείνεται να κανονικοποιούνται οι συχνότητες των όρων ανά έγγραφο για την εκτίμηση της παραμέτρου Poisson, ενώ ο παραδοσιακός πολυωνυμικός ταξινομητής εκτιμά τις παραμέτρους του θεωρώντας όλα τα έγγραφα εκπαίδευσης ως ένα ενιαίο μεγάλο έγγραφο. Επιπλέον, γίνεται αναφορά στην ενίσχυση των βαρών κατηγοριοποίησης για σπάνιες

κατηγορίες, όπου η δημιουργία ενός κατάλληλου πιθανοτικού μοντέλου κατηγοριοποίησης είναι δύσκολη λόγω της έλλειψης εγγράφων εκπαίδευσης. Αντί να δοθεί βάση σε πολλά έγγραφα, γίνεται ενίσχυση στα βάρη κατηγοριοποίησης ορισμένων ενημερωτικών λέξεων, δίνοντάς τους σημαντικό ρόλο στην καθορισμό της ετικέτας του εγγράφου. Πειραματικά αποτελέσματα από δύο συλλογές δείχνουν ότι το προτεινόμενο μοντέλο είναι αρκετά αποδοτικό για την κατασκευή πιθανοτικών ταξινομητών κειμένου με ελάχιστο επιπλέον κόστος σε χρόνο και χώρο, συγκριτικά με τους παραδοσιακούς πολυωνυμικούς ταξινομητές. Η ενίσχυση των βαρών κατηγοριοποίησης είναι ιδιαίτερα αποτελεσματική όταν ο αριθμός των εγγράφων εκπαίδευσης είναι πολύ περιορισμένος. Θεωρείται ότι αυτή η μέθοδος αποτελεί μια καλή εναλλακτική λύση στον παραδοσιακό ταξινομητή Naive Bayes, ο οποίος αποτυγχάνει συχνά λόγω έλλειψης εγγράφων εκπαίδευσης. Αν και ο ταξινομητής Naive Bayes δεν ξεπερνά τις επιδόσεις του σύγχρονου ταξινομητή SVM, μελέτες δείχνουν ότι ο προτεινόμενος ταξινομητής μπορεί να αποδειχθεί ιδιαίτερα χρήσιμος σε πολλές πρακτικές εφαρμογές, όπως ένα σύστημα φιλτραρίσματος σπαμ ή ένα προσαρμοσμένο σύστημα ειδοποιήσεων, λόγω της απλότητας, της αποδοτικότητάς του και της δυνατότητας για σταδιακή μάθηση.

CLASSIFICATION OF EXPLICIT MUSIC CONTENT USING LYRICS AND MUSIC METADATA

Σε έναν κόσμο όπου οι διαδικτυακές πληροφορίες αυξάνονται με ταχύ ρυθμό, η ανάγκη για πιο αποδοτικές μεθόδους αναζήτησης και δημιουργίας μουσικών συλλογών είναι πιο επιτακτική από ποτέ. Εξετάζοντας τις πιο πρόσφατες τάσεις, η εφαρμογή της μηχανικής μάθησης για την αυτοματοποίηση διαφόρων προβλημάτων κατηγοριοποίησης, όπως η ταξινόμηση κατά είδος και διάθεση, έχει δείξει πολύ καλά αποτελέσματα.

Σε αυτή την εργασία, διερευνάται το ζήτημα της κατηγοριοποίησης ρητών μουσικών περιχομένων μέσω μηχανικής μάθησης. Συνδυάζοντας διαφορετικά σύνολα δεδομένων που περιλαμβάνουν στίχους και μεταδεδομένα μουσικής, μεθόδους διανυσματοποίησης και αλγόριθμους, όπως οι Support Vector Machine, Random Forest, k-Nearest Neighbor και Multinomial Naive Bayes, για να δημιουργήσουν 32 διαφορετικές ρυθμίσεις. Οι ρυθμίσεις αυτές αξιολογούνται με τη χρήση καμπυλών ακρίβειας-ανάκλησης.

ΜΕΘΟΔΟΣ:

Αφού καθοριστεί το πρόβλημα, τα απαραίτητα δεδομένα εξάγονται από διάφορες πηγές. Τα δεδομένα προ επεξεργάζονται και διανυσματοποιούνται για τους ταξινομητές κειμένου και στη συνέχεια χωρίζονται σε δεδομένα εκπαίδευσης και δοκιμών. Το σύνολο εκπαίδευσης χρησιμοποιείται για την εκπαίδευση κάθε ταξινομητή, και ο ταξινομητής αξιολογείται χρησιμοποιώντας τα δεδομένα δοκιμών για να σχεδιαστεί μια καμπύλη ακρίβειας - ανάκλησης.

ΣΕΤ ΔΕΔΟΜΕΝΩΝ

Σε αυτήν την εργασία χρησιμοποιούνται δύο διαφορετικά σύνολα δεδομένων. Το ένα περιέχει τους στίχους των τραγουδιών και τις ρητές ετικέτες που χρησιμοποιούνται ως πραγματικά δεδομένα, ενώ το άλλο περιέχει πρόσθετα δεδομένα για κάθε τραγούδι για να εμπλουτίσει τους στίχους.

Στίχοι και Ρητές Ετικέτες:

Οι στίχοι αποκτώνται μέσω του API LyricFind. Το API παρέχει, εκτός από τους στίχους, το όνομα του τραγουδιού, το όνομα του καλλιτέχνη, τη γλώσσα των στίχων και έναν αριθμό αναγνώρισης τραγουδιού που ονομάζεται ISRC, ο οποίος χρησιμοποιείται διεθνώς για τα τραγούδια γενικά. Τα δεδομένα των στίχων συνδέονται με τα χειροκίνητα ελεγμένα τραγούδια από το Soundtrack Your Brand με βάση τον αριθμό ISRC. Αυτό οδηγεί σε σύνολο 27.378 κομματιών.

Μεταδεδομένα Μουσικής:

Προηγούμενες έρευνες έχουν δείξει ότι ο εμπλουτισμός των στίχων με άλλα δεδομένα μπορεί να βελτιώσει τα αποτελέσματα σε διάφορες εργασίες κατηγοριοποίησης κειμένου, στο πλαίσιο της ανάκτησης πληροφοριών μουσικής. Για την κατηγοριοποίηση θεμάτων, έχουν χρησιμοποιηθεί δεδομένα που ανήκουν σε χρήστες, ενώ για την κατηγοριοποίηση κατά είδος έχουν δοκιμαστεί κριτικές πελατών και ετικέτες. Οι στίχοι είναι συχνά σε ποιητική μορφή, χρησιμοποιώντας ομοιοκαταληξίες, και το νόημά τους μπορεί μερικές φορές να παρουσιάζεται έμμεσα μέσω του περιεχομένου, γεγονός που δυσκολεύει έναν αλγόριθμο να κατανοήσει το θέμα του τραγουδιού. Για να τονιστεί το νόημα ενός τραγουδιού, η αρχική ιδέα ήταν να χρησιμοποιηθούν δεδομένα με ανάλυση από χρήστες που περιγράφουν το περιεχόμενο του τραγουδιού. Υπάρχουν πολλές βάσεις δεδομένων που περιέχουν αυτόν τον τύπο πληροφορίας, αλλά λόγω δυσκολιών στην απόκτηση δικαιωμάτων δημοσίευσης και του περιορισμένου χρόνου αυτής της εργασίας, επιλέχθηκε μια διαφορετική κατεύθυνση. Το Spotify προσφέρει μέσω του web API τους ορισμένα μεταδεδομένα για κάθε τραγούδι. Τα μεταδεδομένα περιλαμβάνουν, αλλά δεν περιορίζονται σε συντακτικές πληροφορίες όπως το όνομα του καλλιτέχνη, το άλμπουμ και το έτος, καθώς και ακουστικά χαρακτηριστικά όπως η διάρκεια, το τέμπο, η θετικότητα, η ενέργεια και η τονικότητα.

ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ

Η προεπεξεργασία χρησιμοποιείται για την εξαγωγή των διανυσμάτων χαρακτηριστικών, με σκοπό την κατασκευή και αξιολόγηση του μοντέλου.

Καθαρισμός Δεδομένων:

Το πρώτο βήμα περιλαμβάνει τον καθαρισμό του συνόλου δεδομένων των στίχων. Οι στίχοι που παρέχονται από το LyricFind είναι σε πολλές γλώσσες, οπότε πρέπει να αφαιρεθούν

όλα τα μη αγγλόφωνα τραγούδια. Το σύνολο δεδομένων περιλαμβάνει ένα πεδίο γλώσσας, αλλά σε πολλές εγγραφές λείπει αυτή η πληροφορία, γι' αυτό και η γλώσσα ανιχνεύεται χειροκίνητα για αυτές τις εγγραφές. Επιπλέον, οι εγγραφές με στίχους που είναι μικρότεροι από 100 χαρακτήρες αφαιρούνται, καθώς δεν περιέχουν τους σωστούς ή ολοκληρωμένους στίχους του τραγουδιού. Αυτό περιλαμβάνει, για παράδειγμα, όλες τις εγγραφές με κενό πεδίο στίχων ή τραγούδια που έχουν τη λέξη 'Instrumental' στο πεδίο των στίχων. Μετά από αυτά τα βήματα, απομένουν 25.488 τραγούδια στο σύνολο δεδομένων.

Επιλογή και Μετασχηματισμός Χαρακτηριστικών:

Οι ταξινομητές χρησιμοποιούν δύο διαφορετικά σύνολα δεδομένων: ένα που περιέχει μόνο στίχους και ένα άλλο που περιλαμβάνει στίχους, όνομα καλλιτέχνη, ενέργεια, θετικότητα και έτος. Και τα δύο σύνολα δεδομένων χρησιμοποιούν την ετικέτα ρητών ως ετικέτα ταξινόμησης. Για να συνδυαστούν πολλαπλά χαρακτηριστικά ως μία είσοδος, χρησιμοποιείται το FeatureUnion του Scikit Learn, το οποίο παρέχει τη δυνατότητα συνδυασμού διαφορετικών μεθόδων εξαγωγής χαρακτηριστικών για διάφορα χαρακτηριστικά. Τα χρησιμοποιούμενα χαρακτηριστικά είναι συνδυασμός κειμένου και αριθμών με διαφορετικές απαιτήσεις προεπεξεργασίας.

Εξαγωγή Χαρακτηριστικών:

Για να μετατραπούν οι στίχοι σε κατάλληλη μορφή για ταξινομητές, χρησιμοποιούνται δύο προσεγγίσεις: TF-IDF, που βασίζεται στη συχνότητα λέξεων, και Doc2Vec, που χρησιμοποιεί νευρωνικά δίκτυα για τις σχέσεις μεταξύ λέξεων. Οι παράμετροι της TF-IDF ρυθμίζονται μέσω αναζήτησης σε πλέγμα παραμέτρων, ενώ οι παράμετροι του Doc2Vec βασίζονται σε προτεινόμενες ρυθμίσεις από προηγούμενη έρευνα, λόγω περιορισμών χρόνου.

ΣΥΜΠΕΡΑΣΜΑ:

Η εργασία υλοποιεί και αξιολογεί 32 ρυθμίσεις για την κατηγοριοποίηση ρητού μουσικού περιεχομένου, χρησιμοποιώντας διαφορετικές βάσεις δεδομένων, μεθόδους διανυσματοποίησης και αλγόριθμους. Η ρύθμιση που χρησιμοποιεί τη βάση δεδομένων στίχων, σε συνδυασμό με τη διανυσματοποίηση TF-IDF και τον αλγόριθμο Random Forest, αποδίδει καλύτερα από όλες τις άλλες ρυθμίσεις.

Για μελλοντική εργασία, προτείνεται η συλλογή περισσότερων δεδομένων, όπως η σύνθεση της βάσης LyricFind με τη ρητή βάση δεδομένων. Επίσης, θα μπορούσαν να συμπεριληφθούν περισσότερα χαρακτηριστικά και δεδομένα από χρήστες σχετικά με το περιεχόμενο των τραγουδιών. Ορισμένα στοιχεία στίχων που δεν σχετίζονται με το πραγματικό κείμενο δεν αποκλείστηκαν λόγω περιορισμών χρόνου.

Τέλος, προτείνεται η διερεύνηση μιας πολυδιάστατης προσέγγισης για την ανίχνευση διαφορετικών τύπων ρητής γλώσσας, αν και αυτό απαιτεί τη δημιουργία βάσης αναφοράς για κάθε κατηγορία.

ΚΕΦΑΛΑΙΟ Γ'

Γ' Μοντέλα Βαθιάς Μάθησης

Γ'.1 ΘΕΜΑΤΟΛΟΓΙΑ

Τα 'Μοντέλα Βαθιάς Μάθησης' (Deep Learning Models) είναι μια κατηγορία τεχνητών νευρωνικών δικτύων που χρησιμοποιούνται για την ανάλυση και επεξεργασία δεδομένων με υψηλό επίπεδο πολυπλοκότητας. Αυτά τα μοντέλα αποτελούν βασικό μέρος του πεδίου της μηχανικής μάθησης και πιο συγκεκριμένα της βαθιάς μάθησης, η οποία είναι μια τεχνολογία που εμπνέεται από τον τρόπο που λειτουργεί ο ανθρώπινος εγκέφαλος.

ΒΑΣΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ:

ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΠΟΛΛΩΝ ΕΠΙΠΕΔΩΝ

Τα μοντέλα βαθιάς μάθησης αποτελούνται από πολλά επίπεδα νευρώνων, που ονομάζονται επίσης και 'στρώματα' (layers). Τα πρώτα επίπεδα είναι υπεύθυνα για την αναγνώριση απλών χαρακτηριστικών στα δεδομένα, ενώ τα επόμενα επίπεδα αναλύουν πιο σύνθετα πρότυπα.

ΑΥΤΟΜΑΤΗ ΜΑΘΗΣΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Σε αντίθεση με τις παραδοσιακές μεθόδους μηχανικής μάθησης, όπου τα χαρακτηριστικά των δεδομένων καθορίζονται χειροκίνητα, τα μοντέλα βαθιάς μάθησης μπορούν να μάθουν και να εξάγουν χαρακτηριστικά από τα δεδομένα αυτόματα.

ΕΚΠΑΙΔΕΥΣΗ ΜΕ ΜΕΓΑΛΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ

Τα μοντέλα αυτά χρειάζονται πολύ μεγάλα σύνολα δεδομένων για να εκπαιδευτούν αποτελεσματικά, διότι μόνο έτσι μπορούν να αναγνωρίσουν πολύπλοκα μοτίβα και συσχετίσεις στα δεδομένα.

ΕΦΑΡΜΟΓΕΣ

Χρησιμοποιούνται σε πολλές περιοχές, όπως η αναγνώριση εικόνας και ομιλίας, η επεξεργασία φυσικής γλώσσας και κειμένων, στίχους τραγουδιών, η πρόγνωση τάσεων και πολλά άλλα.

ΠΑΡΑΔΕΙΓΜΑΤΑ ΜΟΝΤΕΛΩΝ ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ:

ΣΥΝΕΛΙΚΤΙΚΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ (CNN)

Χρησιμοποιούνται κυρίως στην αναγνώριση εικόνων και επεξεργασία βίντεο.

ΑΝΑΔΡΑΣΤΙΚΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ (RNN)

Χρησιμοποιούνται για δεδομένα αλληλουχίας, όπως η ανάλυση κειμένου και η επεξεργασία φυσικής γλώσσας.

ΜΕΤΑΣΧΗΜΑΤΙΣΤΕΣ (TRANSFORMERS)

Ιδιαίτερα δημοφιλή σε εφαρμογές επεξεργασίας φυσικής γλώσσας και γλωσσικών μοντέλων.

Τα μοντέλα βαθιάς μάθησης έχουν φέρει επανάσταση σε πολλούς τομείς τεχνολογίας, καθώς μπορούν να επιτύχουν υψηλή ακρίβεια σε προβλήματα που προηγουμένως ήταν πολύ δύσκολο να αντιμετωπιστούν με παραδοσιακές μεθόδους.

Γ'.2 ΕΡΕΥΝΑ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

CORRELATION ANALYSIS BETWEEN HATE SPEECH AND AUDIO COMPONENTS OF SONGS USING SUPERVISED LEARNING

Σε αυτήν την έρευνα, πραγματοποιήθηκαν πειράματα με εκφράσεις μίσους και ηχητική ταξινόμηση χρησιμοποιώντας το Mel Spectrogram, ενώ οι στίχοι των τραγουδιών προήλθαν από τη βάση δεδομένων του Melon. Ο εχθρικός λόγος στους στίχους εντοπίστηκε με τη βοήθεια της Κορεατικής γλώσσας του σώματος και της ομιλίας που χαρακτηρίζεται από εχθρότητα. Ακραίες τιμές απομακρύνθηκαν από τις ετικέτες δεδομένων του εχθρικού λόγου και η διάσταση των δεδομένων μειώθηκε χρησιμοποιώντας τη μέθοδο t-SNE. Στη συνέχεια, η ανάλυση των δεδομένων με μειωμένες διαστάσεις πραγματοποιήθηκε με τον αλγόριθμο K-means, δημιουργώντας τέσσερα συμπλέγματα ετικετών εχθρικού λόγου. Επιπρόσθετα, δημιουργήθηκε μια νέα ετικέτα συνδυάζοντας τις πληροφορίες για τον εχθρικό λόγο με τις πληροφορίες για το μουσικό είδος, και η ηχητική ταξινόμηση έγινε με τη χρήση πέντε προεκπαιδευμένων μοντέλων βαθιάς μάθησης.

Τα αποτελέσματα έδειξαν απόδοση 16,4% στη συνολική ταξινόμηση δεδομένων τραγουδιών και 27,6% στην ταξινόμηση δεδομένων τραγουδιών Hip Hop, υποδεικνύοντας ότι οι ετικέτες εχθρικού λόγου σχετίζονται σημαντικά λιγότερο με το είδος του ήχου.

MUSICNN ΒΙΒΛΙΟΘΗΚΗ:

Η βιβλιοθήκη Musicnn περιέχει ένα σύνολο προ εκπαιδευμένων μουσικών νευρωνικών δικτύων για την προσθήκη ετικετών. Αυτά τα προεκπαιδευμένα μοντέλα βαθιάς μάθησης μπορούν να βελτιστοποιηθούν μαζί με ένα νευρωνικό δίκτυο εξόδου που λειτουργεί ως ταξινομητής, για να εκτελεστεί οποιαδήποτε άλλη μουσική εργασία.

ΜΟΝΤΕΛΑ ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ:

Τα μοντέλα βαθιάς μάθησης που χρησιμοποιούνται στο συγκεκριμένο πείραμα περιλαμβάνουν πέντε τύπους προκαθορισμένων μοντέλων: ResNet18, SqueezeNet, AlexNet, DenseNet, και VGG16, τα οποία παρέχονται από την Pietoch.

i) Το ResNet18 είναι μια αρχιτεκτονική νευρωνικού δικτύου CNN (Convolutional Neural Network) με 18 επίπεδα, που αναπτύχθηκε για να αντιμετωπίσει τις κακές επιδόσεις όταν ο αριθμός των επιπέδων υπερβαίνει ένα συγκεκριμένο όριο σε συμβατικά μοντέλα. Το ResNet χρησιμοποιεί την έννοια της Υπολειμματικής Μάθησης (Residual Learning).

ii) Το AlexNet είναι ένα μοντέλο που αξιοποιεί τη λειτουργία ενεργοποίησης ReLU, την τακτοποίηση εγκατάλειψης και τεχνικές αύξησης δεδομένων, όπως η αναστροφή και η περικοπή εικόνας. Ένα από τα κύρια χαρακτηριστικά του AlexNet είναι η χρήση παράλληλης επεξεργασίας σε δύο GPU, που επιτρέπει ταχύτερους χρόνους εκπαίδευσης και καλύτερη απόδοση.

iii) Το SqueezeNet είναι ένα μοντέλο που σχεδιάστηκε για να έχει όσο το δυνατόν λιγότερες παραμέτρους, διατηρώντας ταυτόχρονα υψηλή απόδοση.

iv) Το VGGNet αναπτύχθηκε για να συνδέει τα επίπεδα στο τέλος του δικτύου, ώστε να εκτελεί την τελική ταξινόμηση. Το κύριο πλεονέκτημά του είναι η απλότητά του, που το καθιστά εύκολο στην προσαρμογή και την εφαρμογή σε ένα ευρύ φάσμα εργασιών. Έχει επίσης χρησιμεύσει ως βάση για πολλές άλλες αρχιτεκτονικές δικτύου, όπως το ResNet και το DenseNet.

v) Το DenseNet είναι ένα παραδοσιακό μοντέλο νευρωνικών δικτύων που στοχεύει στη βελτίωση της ροής πληροφοριών και στη μείωση του αριθμού των παραμέτρων στο δίκτυο.

ΜΕΘΟΔΟΣ:

TAGGING HATE SPEECH TAG

i) Getting the lyrics of the song

Συνολικά ανιχνεύθηκαν 191.245 τραγούδια, εκ των οποίων τα 62,326 επικαλύφθηκαν με το Melon Playlist Continuation σετ δεδομένων. Επομένως, ο αριθμός των τραγουδιών τόσο με το φασματογράφημα όσο και με τους στίχους είναι 62,326.

ii) Translate into Korean

Δεδομένο ότι ο κορμός που χρησιμοποιήθηκε είναι κατασκευασμένος για Κορεάτικα, πρέπει να μεταφραστεί η ξένη γλώσσα που περιλαμβάνεται στους Κορεάτικους στίχους. Για να μεταφραστεί μεγάλος όγκος στίχων, χρησιμοποιήθηκε η λειτουργία GoogleTranslate.

iii) Tagging

Αυτή είναι η ταξινόμηση μεροληψίας και μίσους που χρησιμοποιεί, τα KoELECTRA και τον Κορεάτικο εχθρικό λόγο, σύνολα δεδομένων, όπου προσθέτει ετικέτες μεροληψίας (Gender, Other, None) και μίσους (Hate, Offensive, None) στις εισαγόμενες προτάσεις. Δεδομένου ότι αυτό το μοντέλο ήταν ένας τρόπος προσθήκης ετικετών στα σχόλια, η μελέτη χρησιμοποίησε μία μέθοδο επισήμανσης κάθε γραμμής στίχων τα οποία και στη συνέχεια τα συγκέντρωσε μαζί. Έτσι, οι ετικέτες που δίνονται σε κάθε τραγούδι έχουν μορφή διανυσμάτων της αναλογίας μεροληψίας και μίσους.

AUDIO DATA SELECTION

i) Remove bad data

Αφαιρέθηκαν τραγούδια που δεν έχουν Genre Code από τα μεταδεδομένα του Arena. Επίσης, αφαιρέθηκαν δεδομένα όπου η διάρκεια του τραγουδιού δεν ήταν σταθερή.

ii) Hate tag-based data clustering

Οι τιμές που αποδίδονται σε έξι στήλες μιας ετικέτας μίσους ήταν γύρω στο δεύτερο δεκαδικό ψηφίο. Έτσι, χρησιμοποιήθηκε το Isolation Forest για αυτές τις ετικέτες μίσους για την εξάλειψη των ακραίων δεδομένων.

ΣΥΜΠΕΡΑΣΜΑ:

Πειράματα έδειξαν ότι οι ηχητικές αναπαραστάσεις και οι ετικέτες μεροληψίας και εχθρικού λόγου δεν έχουν μεγάλη συσχέτιση. Ωστόσο, αξιολογήθηκε η απόδοση της ετικέτας του εχθρικού λόγου - μεροληψίας ταξινόμησης για μεμονωμένα είδη και διαπιστώθηκε ότι η απόδοση βελτιώθηκε σε σύγκριση με την ταξινόμηση για όλο το είδος. Γι' αυτό τον λόγο, κρίνεται η βελτίωση της μεθόδου επισήμανσης ετικετών μεροληψίας και εχθρικού λόγου στίχων και η μέθοδος ομαδοποίησης των παραπάνω θα βελτίωναν την απόδοση του μοντέλου.

DETECTING EXPLICIT LYRICS: A CASE STUDY IN ITALIAN MUSIC

Διερευνήθηκε η αυτόματη ανίχνευση ακατάλληλων στίχων για ιταλικά τραγούδια, επεκτείνοντας τις αναλύσεις που έχουν γίνει σε άλλες γλώσσες. Για τη σύγκριση των διαφόρων συστημάτων, χρησιμοποιήθηκε ένα νέο σύνολο δεδομένων, το οποίο περιλαμβάνει περίπου 34.000 τραγούδια με ετικέτες που υποδεικνύουν ακατάλληλο περιεχόμενο. Η αξιολόγηση δείχνει ότι, σε αυτό το σύνολο δεδομένων, οι περισσότεροι από τους ταξινομητές που βασίζονται σε νευρωνικά γλωσσικά μοντέλα αποδίδουν πολύ καλύτερα από τις μη νευρωνικές μεθόδους. Παρέχονται επίσης πρόσθετες αναλύσεις, όπως μια ποιοτική αξιολόγηση των προβλέψεων των ταξινομητών, μια αξιολόγηση της απόδοσης του καλύτερου ταξινομητή σε σενάριο εκμάθησης με λίγα δείγματα και την επίδραση της εξισορρόπησης του συνόλου δεδομένων.

ΤΑΞΙΝΟΜΗΤΕΣ ΤΩΝ NLMs:

Τα Νευρωνικά Γλωσσικά Μοντέλα (Neural Language Models, NLMs) είναι μεγάλης κλίμακας προεκπαιδευμένα βαθιά μοντέλα αναπαράστασης γλώσσας που μπορούν να χρησιμοποιηθούν για την εκτέλεση πολλών εργασιών κατανόησης φυσικής γλώσσας. Η ανάπτυξή τους δυνατοποιείται από τους Μετασχηματιστές (Transformers), νευρωνικές αρχιτεκτονικές Seq2Seq που χρησιμοποιούν την προσοχή για να αποτυπώσουν τις εξαρτήσεις μεταξύ των ακολουθιών εισόδου και εξόδου. Μόλις προ εκπαιδευτούν (σε μια γενική εργασία όπως η μοντελοποίηση μάσκας γλώσσας), ένα NLM μπορεί να προσαρμοστεί (fine-tuned) στην συγκεκριμένη εργασία επεξεργασίας φυσικής γλώσσας που εξετάζεται: αυτό συνίσταται στην προσαρμογή και εκπαίδευση του τελευταίου επιπέδου εξόδου του δικτύου μετασχηματιστή, διατηρώντας έτσι όλες τις σημασιολογικές πληροφορίες και τη γνώση της γλώσσας που έχουν αποτυπωθεί στα προηγούμενα επίπεδα του δικτύου. Παρακάτω, γίνεται εξέταση και σύγκριση μερικών NLMs, κατάλληλα για Ιταλικό κείμενο:

Multilingual-BERT

Αναπτύχθηκε από τη Google, αξιοποιεί μια βαθιά διδιάκριτη αναπαράσταση που έχει μάθει από ασημειωτό κείμενο από τη Wikipedia σε πολλές γλώσσες. Πρόκειται για ένα μοντέλο μονής γλώσσας προεκπαιδευμένο από μονογλωσσικά σώματα κειμένου σε 104 γλώσσες, συμπεριλαμβανομένων και των ιταλικών.

Italian-BERT

Αναπτύχθηκε από την ομάδα MDZ Digital Library. Είναι ένα μοντέλο BERT για τα ιταλικά που εκπαιδεύτηκε στην ιταλική Wikipedia και σε διάφορα κείμενα από τη συλλογή σώματος κειμένων OPUS.

aIBERTo

Αναπτύχθηκε από το Πανεπιστήμιο της Μπάρι, είναι ένα γλωσσικό μοντέλο BERT για τα ιταλικά, εκπαιδευμένο στην ιταλική γλώσσα που χρησιμοποιείται στα κοινωνικά δίκτυα (Twitter).

UmBERTo-Commoncrawl & UmBERTo-Wikipedia

Αναπτύχθηκαν από την Musixmatch AI, είναι δύο μοντέλα γλωσσικών RoBERTa που εκπαιδεύτηκαν σε διαφορετικά ιταλικά σώματα κειμένων.

Το πρώτο είναι εκπαιδευμένο στο Commoncrawl ITA αξιοποιώντας το OSCAR (Open Super-large Crawled ALMANACH corpus).

Το δεύτερο είναι εκπαιδευμένο σε ένα σύνολο κειμένων εξαγόμενων από την ιταλική Wikipedia.

GilBERTo

Είναι ένα ιταλικό προεκπαιδευμένο γλωσσικό μοντέλο βασισμένο σε RoBERTa, εκμεταλλευόμενο την προσέγγιση της τοκενοποίησης κειμένου CamemBERT. Εκπαιδεύτηκε στο OSCAR.

Για την εφαρμογή των NLMs για την ανίχνευση ακατάλληλων στίχων, αρχικά οι στίχοι πρέπει να υποστούν κατάλληλη τοκενοποίηση δηλαδή να τεκμηριωθούν κατάλληλα. Στη συνέχεια, προστίθεται ένα επίπεδο ταξινόμησης ακολουθιών με ενεργοποίηση σιγμοειδούς πάνω από κάθε προεκπαιδευμένο μοντέλο και γίνεται λεπτομερής προσαρμογή για τη συγκεκριμένη εργασία.

ΣΕΤ ΔΕΔΟΜΕΝΩΝ ΙΤΑΛΙΚΩΝ ΣΤΙΧΩΝ:

Για τη δημιουργία του, χρησιμοποιήθηκε το LyricWiki και το Spotify. Μετά από επεξεργασία των στίχων και την εξαγωγή των απαραίτητων πληροφοριών από το LyricWiki, χρησιμοποιήθηκε το Spotify API για την ανάκτηση των μεταδεδομένων πληροφοριών και την τιμή "explicit". Όπως και ήταν αναμενόμενο, το σύνολο δεδομένων είναι μη ισορροπημένο, καθώς οι στίχοι που δεν περιέχουν ακατάλληλο περιεχόμενο είναι πολύ περισσότεροι από τους ακατάλληλους.

ΣΥΜΠΕΡΑΣΜΑ:

Μελετήθηκε το πρόβλημα της ανίχνευσης εκφράσεων στίχων τραγουδιών στα ιταλικά που θα μπορούσαν να θεωρηθούν προσβλητικοί ή ακατάλληλοι για παιδιά. Εμπνευσμένοι από προηγούμενες μελέτες σε αγγλικά και κορεάτικα τραγούδια, όπου το πρόβλημα είχε διατυπωθεί ως δυαδική κατηγοριοποίηση, αξιολογήθηκε η απόδοση πολλών ταξινομητών μηχανικής μάθησης. Αυτή η αξιολόγηση έγινε σε ένα νέο σύνολο δεδομένων ιταλικών στίχων, το οποίο αποτελείται από 34 χιλιάδες τραγούδια με επισημασμένες πληροφορίες για ακατάλληλο περιεχόμενο.

Η αξιολόγηση έδειξε ότι οι ταξινομητές νευρωνικών μοντέλων γλώσσας επιδοτούν σημαντικά καλύτερα από άλλες μεθόδους (ο καλύτερος ταξινομητής NLM βελτίωσε την απόδοση κατά 0.040 σε σύγκριση με μη-NLM ταξινομητή), με τα UMBERTO-CC, ALBERTO και ITALIAN-BERT να επιτυγχάνουν τις υψηλότερες βαθμολογίες.

A LITERATURE REVIEW OF TEXTUAL HATE SPEECH DETECTION METHODS AND DATASETS

Η ρητορική μίσους είναι ένα περίπλοκο και επιβλαβές περιεχόμενο που στοχεύει άτομα ή ομάδες. Αυτή η μελέτη ανασκοπεί συστήματα ανίχνευσης ρητορικής μίσους, επισημαίνοντας τα βασικά σύνολα δεδομένων και μοντέλα μηχανικής μάθησης. Τα ευρήματα δείχνουν ότι πολλά σύνολα δεδομένων είναι μικρά και αναξιόπιστα, ενώ οι πιο επιτυχημένες μέθοδοι συνδυάζουν πολλαπλά μοντέλα βαθιάς μάθησης. Η μελέτη παρέχει χρήσιμες γνώσεις για τη ρητορική μίσους και κατευθύνσεις για μελλοντική έρευνα.

ΜΕΘΟΔΟΙ ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ:

Τα CNNs είναι νευρωνικά δίκτυα κυρίως για ανάλυση εικόνας αλλά και για ταξινόμηση κειμένου, όπως ανίχνευση μίσους.

Τα RNNs χρησιμοποιούν μνήμη για να θυμούνται την είσοδο τους, με αλγόριθμους όπως LSTMs και GRUs.

Τα BERTs είναι μοντέλα μετασχηματιστών για επεξεργασία φυσικής γλώσσας, αποδεικνύοντας την αποτελεσματικότητά τους στην ανίχνευση εχθρικού λόγου.

Τα συνελκτικά νευρωνικά δίκτυα (CNNs) αποτελούν μια κατηγορία από τα τεχνητά νευρωνικά δίκτυα που αρχικά χρησιμοποιούνταν για την ανάλυση οπτικών εικόνων, αλλά έχουν προσαρμοστεί επιτυχώς και για την ταξινόμηση κειμένου, συμπεριλαμβανομένης της

ανίχνευσης εχθρικού λόγου. Χρησιμοποιούν φίλτρα και στρώματα pooling για την εξαγωγή χαρακτηριστικών από τα δεδομένα εισόδου.

Τα αναδρομικά νευρωνικά δίκτυα (RNNs) αποτελούν ένα άλλο είδος νευρωνικών δικτύων που χρησιμοποιούν εσωτερική μνήμη για να θυμούνται την προηγούμενη είσοδο τους. Αλγόριθμοι όπως τα Long Short-Term Memory networks (LSTMs) και οι Gated Recurrent Units (GRUs) είναι δημοφιλείς επιλογές στα RNNs λόγω της ικανότητάς τους να διαχειρίζονται το πρόβλημα της εξαφάνισης της κλίσης που αντιμετωπίζεται στα παραδοσιακά RNNs.

Οι αναπαραστάσεις διπλής κατεύθυνσης από μετασχηματιστές (BERTs) αποτελούν προηγμένη τεχνική μηχανικής μάθησης που χρησιμοποιείται για την επεξεργασία φυσικής γλώσσας. Αρχικά προ εκπαιδεύονται σε μεγάλα σύνολα δεδομένων κειμένου και στη συνέχεια εφαρμόζονται σε διάφορες εργασίες ανάλυσης, συμπεριλαμβανομένης της ανίχνευσης εχθρικού λόγου.

ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ:

Οι συλλογές δεδομένων που χρησιμοποιούνται για την ανάλυση ρητορικής μίσους καλύπτουν πολλές διαφορετικές κατηγορίες. Ωστόσο, πολλές φορές τα παραδείγματα σε αυτά τα σύνολα δεδομένων μπορεί να είναι ασαφή, όπως συμβαίνει στο σύνολο δεδομένων του Waseem ή σε ιεραρχικά σύνολα όπως το σύνολο δεδομένων των Basile. Επιπλέον, η ποιότητα των συνόλων δεδομένων παραμένει πρόβλημα εξαιτίας της έλλειψης τακτικών ενημερώσεων για να αντανakλούν τις νέες φράσεις ή συντομογραφίες που χρησιμοποιούν οι χρήστες του Twitter. Επιπλέον, περίπου το 60% των δημιουργών συνόλων δεδομένων συμφωνούν μεταξύ τους, γεγονός που μπορεί να επηρεάζει την αντικειμενικότητα και την αξιοπιστία της αναγνώρισης εχθρικού λόγου. Οι Koco και οι συνεργάτες του έχουν δείξει ότι ο διαχωρισμός των ομάδων σχολιαστών επηρεάζει σημαντικά την απόδοση των συστημάτων ανίχνευσης μίσους. Επιπλέον, η συναίνεση της ομάδας σχολιαστών επηρεάζει την ποιότητα της αναγνώρισης, καθώς η ομάδα αναγνώρισης μπορεί να εισάγει προκαταλήψεις στο σύνολο δεδομένων.

ΣΥΜΠΕΡΑΣΜΑ:

Νέα σύνολα δεδομένων ρητορικής μίσους αναπτύσσονται συνεχώς, αλλά πολλά είναι μικρά ή αναξιόπιστα λόγω του τρόπου συλλογής και επισημάνσεων. Η έλλειψη γλωσσικής ποικιλίας και η διαφορετική προέλευση των δεδομένων δυσκολεύουν τη σύγκριση των μοντέλων ανίχνευσης ρητορικής μίσους. Οι προκλήσεις περιλαμβάνουν την αρχιτεκτονική των μοντέλων μηχανικής μάθησης και την έλλειψη συμφωνίας στον ορισμό της ρητορικής μίσους. Η βιβλιογραφία δείχνει ότι απαιτούνται βελτιώσεις στα σύνολα δεδομένων και στην επιλογή χαρακτηριστικών για την ανίχνευση ρητορικής μίσους.

Σύμφωνα με έρευνες, οι πιο υποσχόμενοι αλγόριθμοι στις μεθόδους βαθιάς μάθησης είναι τα CNN και LSTM. Ωστόσο, η απόδοση αυτών των προσεγγίσεων ποικίλλει και η πλειονότητά τους επικεντρώνεται σε ένα δυαδικό πρόβλημα ανίχνευσης μίσους (μίσος/όχι μίσος), ενώ οι πολυετικέτες ταξινομήσεις εξετάζονται σπάνια. Αυτό μπορεί να οφείλεται στην έλλειψη συνόλων δεδομένων ή στις δυσκολίες ανίχνευσης ρητορικής μίσους. Παρόλα αυτά, τα CNN και LSTM έχουν αποδειχθεί αποτελεσματικά σε πολλές δοκιμές.

HATE SPEECH DETECTION IN INDONESIAN SONG LYRICS USING BERT

Η μελέτη αυτή διεξήχθη με στόχο την ανίχνευση εχθρικού λόγου στους στίχους τραγουδιών. Το υπόβαθρο της μελέτης αυτής προέρχεται από τις πολυάριθμες αρνητικές επιπτώσεις που προκαλούνται από τη ρητορική μίσους. Σύμφωνα με τις προηγούμενες εργασίες, οι αλγόριθμοι βαθιάς μάθησης, όπως η χρήση του προκαταρκτικά εκπαιδευμένου μοντέλου BERT για ανάλυση συναισθήματος και του R-CNN ως ταξινομητή, έχουν δείξει ελπίδα στην αντιμετώπιση αυτών των ζητημάτων. Επομένως, αποφασίστηκε να χρησιμοποιηθεί το προκαταρκτικά εκπαιδευμένο μοντέλο BERT για ανάλυση συναισθήματος και το R-CNN ως ταξινομητή στη συγκεκριμένη μελέτη. Το σύνολο δεδομένων που χρησιμοποιήθηκε σε αυτή την έρευνα συλλέχθηκε από το σύνολο δεδομένων Twitter στο Kaggle και επίσης χρησιμοποιήθηκε το API του Spotify. Το αποτέλεσμα υποδηλώνει ότι περισσότερα συναφή δεδομένα, όπως σύνολα δεδομένων που αποτελούνται από ποίηση ή προτάσεις με έμμεσες σημασίες, θα είναι επωφελή για την ενίσχυση της ανίχνευσης του εχθρικού λόγου στους στίχους τραγουδιών.

ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ:

Αυτή η έρευνα αξιοποιεί δύο διαφορετικά σύνολα δεδομένων:

Το σύνολο δεδομένων Kaggle με τίτλο "Indonesian Abusive and Hate Speech Twitter Text" και ένα άλλο που αποκτήθηκε από το Spotify API. Το σύνολο δεδομένων Kaggle περιλαμβάνει επισημασμένα tweets, με ετικέτες 1 (HS) για ρητορική μίσους και 0 (Non-HS) για μη ρητορική μίσους, συνολικά 13.023 καταχωρήσεις δεδομένων. Τα δεδομένα τραγουδιών περιλαμβάνουν 100 ινδονησιακά τραγούδια, με τα ονόματα και τους στίχους τους, τα οποία επισημάνθηκαν χειροκίνητα για να δημιουργήσουν ένα σώμα ανάλυσης συναισθήματος.

ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ:

Η προεπεξεργασία δεδομένων είναι μια τεχνική που χρησιμοποιείται για την επεξεργασία των συλλεγμένων δεδομένων ώστε να προσαρμοστούν για το μοντέλο εισόδου στα στάδια της εκπαίδευσης ή της δοκιμής. Αυτή η διαδικασία περιλαμβάνει διάφορα στάδια:

LOWERCASING

Η διαδικασία μετατροπής σε πεζά γράμματα πραγματοποιείται αλλάζοντας το κείμενο σε πεζή μορφή χωρίς να αλλάζει η δομή ή το περιεχόμενό του. Αυτή η μετατροπή εφαρμόζεται για να μειωθεί ο αριθμός των μοναδικών λέξεων που περιέχονται στο σύνολο δεδομένων.

REMOVE NON ALPHABETIC CHARACTER

Οι αριθμοί, τα σημεία στίξης και οι ειδικοί χαρακτήρες αφαιρούνται από το κείμενο για να το καθαρίσουν και να μειώσουν τον θόρυβο. Αυτή η διαδικασία συμβάλλει επίσης στη μείωση της κατανάλωσης μνήμης και επιταχύνει την εκμάθηση κατά την επεξεργασία των δεδομένων στο στάδιο της εκπαίδευσης.

REMOVE STOPWORD

Αφαίρεση κοινών λέξεων που δεν προσφέρουν σημαντικό νόημα στο κείμενο. Η αφαίρεση αυτών των λέξεων (stopwords) καθιστά το κείμενο πιο συνοπτικό και επικεντρωμένο στο ουσιαστικό περιεχόμενο. Η διαδικασία φιλτραρίσματος των δεδομένων εκτελείται χρησιμοποιώντας το σύνολο δεδομένων stopwords από την ίδια πηγή με το σύνολο δεδομένων Twitter.

STEMMING

Η διαδικασία της ριζοποίησης (stemming) επιτυγχάνεται με την αλλαγή των λέξεων στην βασική ή τη ριζική τους μορφή. Σκοπός της ριζοποίησης είναι η κανονικοποίηση λέξεων που έχουν την ίδια σημασία αλλά βρίσκονται σε διαφορετικές μορφές.

ΜΕΘΟΔΟΣ:

Αυτή η μελέτη εστιάζει στην ανάλυση συναισθήματος του εχθρικού λόγου σε τραγούδια, η οποία αναλύεται χρησιμοποιώντας το βελτιστοποιημένο μοντέλο IndoBERT.

INDOBERT

Το 'BERT' σημαίνει Bidirectional Encoder Representations from Transformers. Είναι ένα μοντέλο γλώσσας που έχει εκπαιδευτεί προκαταρκτικά και προτάθηκε από την Google το

2018. Το μοντέλο εκπαιδεύεται με τη χρήση ενός μεγάλου αριθμού μη επισημασμένων κειμένων, χωρίς επίβλεψη.

RCNN

Το Recurrent Convolutional Neural Network 'RCNN' είναι ένα μοντέλο που συνδυάζει αναδρομικές και συνελκτικές δομές. Ο συνδυασμός του RCNN με το IndoBERT για την ενσωμάτωση λέξεων βασίζεται σε προηγούμενες έρευνες που είχαν βελτιωμένα αποτελέσματα με τη χρήση του συνδυασμού IndoBERT και RCNN ως ταξινομητή.

ΣΥΜΠΕΡΑΣΜΑ:

Αυτή η μελέτη βελτίωσε το προεκπαιδευμένο μοντέλο IndoBERT με τη χρήση του RCNN για την αναγνώριση εχθρικού λόγου στο σύνολο δεδομένων του Twitter. Τα τελικά αποτελέσματα της έρευνας έδειξαν ότι η χρήση του IndoBERT και του RCNN παρήγαγε καλύτερα αποτελέσματα με ακρίβεια 77,78% σε σύγκριση με τη χρήση μόνο του IndoBERT.

ON EXPLOITING TRANSFORMERS FOR DETECTING EXPLICIT SONG LYRICS

Η αξιολόγηση εάν οι στίχοι ενός τραγουδιού ενδέχεται να είναι επιβλαβείς ή ακατάλληλοι για παιδιά είναι κρίσιμη για την αποφυγή της αναπαραγωγής τραγουδιών με ακατάλληλο περιεχόμενο. Αυτό το ζήτημα μπορεί να αντιμετωπιστεί υπολογιστικά ως πρόβλημα δυαδικής ταξινόμησης, και τα τελευταία χρόνια έχουν εφαρμοστεί διάφορες μέθοδοι μηχανικής μάθησης για την αυτόματη εκτέλεση αυτής της εργασίας. Στην παρούσα εργασία, εξετάζεται η αυτόματη ανίχνευση ρητών στίχων τραγουδιών χρησιμοποιώντας γλωσσικά μοντέλα βασισμένα σε μετασχηματιστές (transformers). Αυτά τα μοντέλα είναι μεγάλες γλωσσικές αναπαραστάσεις που έχουν δημιουργηθεί μη επιβλεπόμενα από εκτενή κείμενα και μπορούν να προσαρμοστούν σε διάφορα καθήκοντα επεξεργασίας φυσικής γλώσσας, όπως η ταξινόμηση κειμένων. Αξιολογείται η απόδοση διαφόρων ταξινομητών, βασισμένων σε transformers, σε ένα σύνολο δεδομένων που περιλαμβάνει πάνω από 800.000 στίχους με επισημάνσεις για ρητό περιεχόμενο. Τα αποτελέσματα δείχνουν ότι αν και οι ταξινομητές που χρησιμοποιούν αυτά τα ισχυρά εργαλεία επιτυγχάνουν υψηλή απόδοση, δεν ξεπερνούν ελαφρύτερες και λιγότερο υπολογιστικά απαιτητικές μεθόδους.

ΜΕΘΟΔΟΣ:

ΔΕΔΟΜΕΝΑ ΕΚΠΑΙΔΕΥΣΗΣ

Η έρευνα χρησιμοποιεί ένα μεγάλο σύνολο δεδομένων με πάνω από 800.000 στίχους τραγουδιών, οι οποίοι έχουν προ επισημανθεί ως ρητοί ή μη ρητοί. Αυτή η προ επισημάνση γίνεται πιθανόν από πλατφόρμες streaming που επισημαίνουν το ρητό περιεχόμενο.

FINE-TUNING TRANSFORMERS

Τα γλωσσικά μοντέλα transformers προσαρμόζονται σε αυτή την εργασία μέσω μιας διαδικασίας που ονομάζεται fine-tuning. Η διαδικασία αυτή περιλαμβάνει την περαιτέρω εκπαίδευση του μοντέλου πάνω στο συγκεκριμένο σύνολο δεδομένων, ώστε να βελτιστοποιηθεί για την αναγνώριση του ρητού περιεχομένου.

ΑΞΙΟΛΟΓΗΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Η απόδοση των μοντέλων αξιολογείται χρησιμοποιώντας μετρικές όπως η ακρίβεια (accuracy) και η F1-score. Τα αποτελέσματα συγκρίνονται με πιο απλά μοντέλα, όπως η λογιστική παλινδρόμηση και οι Μηχανές Υποστήριξης Διανυσμάτων (SVM). Ένα σημαντικό εύρημα είναι ότι, παρά την πολυπλοκότητα των transformers, δεν υπερέχουν δραματικά των ελαφρύτερων μοντέλων σε όρους ακρίβειας.

FEW-SHOT LEARNING

Μια ακόμα ενδιαφέρουσα προσέγγιση που εξετάζεται είναι η χρήση της μάθησης με λίγα δεδομένα (few-shot learning). Εδώ, το μοντέλο εκπαιδεύεται με έναν πολύ μικρότερο αριθμό δειγμάτων για να εξεταστεί η απόδοσή του σε σενάρια όπου τα δεδομένα είναι περιορισμένα.

ΣΥΜΠΕΡΑΣΜΑ:

Η έρευνα καταλήγει ότι ενώ οι μετασχηματιστές (transformers) επιτυγχάνουν υψηλή απόδοση, τα απλούστερα μοντέλα μπορεί να είναι πιο κατάλληλα για πραγματικές εφαρμογές λόγω της χαμηλότερης κατανάλωσης υπολογιστικών πόρων. Τα αποτελέσματα προτείνουν επίσης ότι η μέθοδος αυτή μπορεί να εφαρμοστεί σε πλατφόρμες όπως το Spotify για το φιλτράρισμα περιεχομένου σε πραγματικό χρόνο.

ΚΕΦΑΛΑΙΟ Δ'

Δ' Συμπεράσματα

Η σύγκριση των παραδοσιακών μεθόδων με τα μοντέλα βαθιάς μάθησης στην ανάλυση και αναγνώριση εχθρικού λόγου σε στίχους τραγουδιών, αποκαλύπτει σημαντικές διαφορές στις επιδόσεις, τις δυνατότητες και τις προκλήσεις της κάθε προσέγγισης. Στη συνέχεια, αναφέρονται οι κύριες πτυχές της σύγκρισης και τα αντίστοιχα συμπεράσματα τους.

ΑΚΡΙΒΕΙΑ ΚΑΙ ΑΠΟΔΟΣΗ

Οι παραδοσιακές μέθοδοι, όπως τα Naive Bayes, SVM ή Random Forest κ.α, προσφέρουν καλή ακρίβεια σε μικρότερα και πιο απλά δεδομένα. Οι τεχνικές αυτές τείνουν να είναι ελαφρές και γρήγορες, αλλά συχνά δεν καταφέρνουν να κατανοήσουν πλήρως το πλαίσιο των λέξεων ή τη σημασία που προκύπτει από πιο σύνθετες σχέσεις μεταξύ λέξεων. Η απόδοσή τους μπορεί να είναι επαρκής σε περιπτώσεις όπου ο λόγος είναι πιο ευθύς και απλός. Τα μοντέλα βαθιάς μάθησης, όπως τα LSTM, CNN, BERT κ.α, υπερτερούν σημαντικά όσον αφορά την κατανόηση του πλαισίου και των σύνθετων σχέσεων μεταξύ λέξεων. Τα μοντέλα αυτά βασίζονται σε μεγάλο όγκο δεδομένων για εκπαίδευση, κάτι που τα καθιστά ιδιαίτερα αποτελεσματικά στην κατανόηση του πολυσύνθετου και πολυεπίπεδου λόγου που συναντάται σε στίχους τραγουδιών, όπου η χρήση μεταφορών και κρυφών νοημάτων είναι συχνή. Αυτό οδηγεί σε υψηλότερη ακρίβεια και απόδοση στην αναγνώριση εχθρικού λόγου.

ΑΝΑΛΥΣΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ

Για τις παραδοσιακές μεθόδους, η ανάλυση βασίζεται συνήθως σε απλές συσχετίσεις μεταξύ των λέξεων. Χρησιμοποιούν τεχνικές όπως το Bag of Words (BOW) ή TF-IDF, που βασίζονται μόνο στη συχνότητα λέξεων και δεν λαμβάνουν υπόψη τη σημασία του συμφραζομένου. Αυτό καθιστά δύσκολη την κατανόηση της πολυσημίας και των σύνθετων νοημάτων που εμφανίζονται συχνά στη μουσική. Από την άλλη, για τα μοντέλα βαθιάς μάθησης και συγκεκριμένα τα μοντέλα όπως το BERT και το GPT έχουν σχεδιαστεί ειδικά για την κατανόηση του συμφραζομένου. Αυτά τα μοντέλα βασίζονται στην αναγνώριση των σχέσεων μεταξύ των λέξεων μέσα σε ένα μεγαλύτερο κείμενο και μπορούν να κατανοήσουν τον εχθρικό λόγο που προκύπτει έμμεσα ή μέσα από το συνολικό νόημα ενός στίχου. Επιτρέπουν την ανάλυση της σημασίας λέξεων σε διαφορετικά συμφραζόμενα, κάτι που τα κάνει πολύ πιο ισχυρά στην αναγνώριση λεπτών μορφών εχθρικού λόγου.

ΑΠΑΙΤΗΣΕΙΣ ΠΟΡΩΝ

Οι παραδοσιακές μέθοδοι χρειάζονται μικρότερη υπολογιστική ισχύ και είναι λιγότερο απαιτητικές όσον αφορά τους πόρους. Τα αλγοριθμικά μοντέλα μπορούν να εκπαιδευτούν σχετικά γρήγορα και να χρησιμοποιηθούν σε εφαρμογές με περιορισμένο υπολογιστικό χρόνο και χωρητικότητα. Αντίθετα, τα μοντέλα βαθιάς μάθησης απαιτούν μεγάλη επεξεργαστική ισχύ και εξειδικευμένο υλικό για να εκπαιδευτούν και να λειτουργήσουν αποδοτικά. Αυτό σημαίνει ότι είναι πιο αργά στην εκπαίδευση και πιο δαπανηρά όσον αφορά την ενέργεια και τον χρόνο. Ωστόσο, μετά την αρχική τους εκπαίδευση, η απόδοσή τους είναι συνήθως ανώτερη σε απαιτητικές εφαρμογές.

ΠΡΟΣΑΡΜΟΣΤΙΚΟΤΗΤΑ

Τα παραδοσιακά μοντέλα είναι ευκολότερα να επεκταθούν σε νέες κατηγορίες δεδομένων και μπορούν να προσαρμοστούν γρήγορα με μικρότερη ποσότητα δεδομένων εκπαίδευσης. Είναι κατάλληλα για μικρότερες εφαρμογές ή για περιπτώσεις όπου υπάρχει μικρός αριθμός διαθέσιμων δεδομένων. Τα μοντέλα βαθιάς μάθησης χρειάζονται μεγάλα σύνολα δεδομένων για εκπαίδευση και είναι πιο δύσκολα στην προσαρμογή σε νέα δεδομένα χωρίς επανεκπαίδευση με μεγάλο αριθμό παραδειγμάτων. Παράλληλα, όμως, τα μοντέλα αυτά, μόλις εκπαιδευτούν, είναι εξαιρετικά ευέλικτα και μπορούν να αναγνωρίζουν πιο λεπτές αποχρώσεις και μοτίβα που δεν θα μπορούσαν να καταλάβουν οι παραδοσιακές μέθοδοι.

ΔΙΑΧΕΙΡΙΣΗ ΠΟΛΥΠΛΟΚΟΤΗΤΑΣ ΓΛΩΣΣΑΣ

Τα παραδοσιακά μοντέλα έχουν περιορισμούς στη διαχείριση της πολυπλοκότητας της γλώσσας, ιδιαίτερα σε κείμενα όπως οι στίχοι τραγουδιών που μπορεί να περιέχουν μεταφορές, ασαφή νοήματα ή λογοπαίγνια. Δεν έχουν τη δυνατότητα να αντιληφθούν τις σύνθετες γλωσσικές δομές και πολλές φορές οδηγούν σε λανθασμένα συμπεράσματα. Τα μοντέλα βαθιάς μάθησης, ειδικά τα 'Transformer-based μοντέλα' όπως το BERT ή το GPT, είναι σχεδιασμένα για να κατανοούν τη φυσική γλώσσα με μεγάλο βαθμό ακρίβειας. Μπορούν να αντιμετωπίσουν τις πολυπλοκότητες της γλώσσας, όπως την πολυσημία και τις μεταφορικές έννοιες που συναντώνται συχνά στους στίχους τραγουδιών.

Η σύγκριση των παραδοσιακών μεθόδων και των μοντέλων βαθιάς μάθησης αναδεικνύει τα εξής βασικά συμπεράσματα:

Οι παραδοσιακές μέθοδοι είναι πιο γρήγορες και λιγότερο απαιτητικές σε πόρους, καθιστώντας τις κατάλληλες για εφαρμογές με μικρά δεδομένα ή περιορισμένο υπολογιστικό χρόνο. Επίσης, μπορεί να είναι αρκετά ικανοποιητικές για απλούστερες εφαρμογές, αλλά οι

σύγχρονες απαιτήσεις, όπως η κατανόηση του εχθρικού λόγου σε περιβάλλοντα με μεγάλη γλωσσική ποικιλία, καθιστούν αναγκαία τη χρήση βαθιάς μάθησης για καλύτερα αποτελέσματα.

Τα μοντέλα βαθιάς μάθησης προσφέρουν καλύτερη απόδοση και ακρίβεια, ειδικά σε δεδομένα που περιλαμβάνουν σύνθετα γλωσσικά μοτίβα, όπως στην προκειμένη περίπτωση, οι στίχοι τραγουδιών. Παρέχουν πολύ καλύτερη κατανόηση των συμφραζόμενων και της πολυπλοκότητας της γλώσσας. Γενικά, είναι πιο δύσκολα στην εκπαίδευση και απαιτούν μεγαλύτερα σύνολα δεδομένων, όμως αποδίδουν εξαιρετικά σε μεγάλης κλίμακας εφαρμογές.

ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ:

Η μελλοντική έρευνα για την αναγνώριση εχθρικού λόγου στους στίχους τραγουδιών πρέπει να επικεντρωθεί στη βελτίωση της κατανόησης της γλώσσας σε συμφραζόμενα, στην ηθική χρήση των αλγορίθμων, στην ανάπτυξη πολυγλωσσικών και πολυτροπικών μοντέλων, και στην ενσωμάτωση προσαρμοστικών τεχνικών μάθησης.

Τα Νευρωνικά Μοντέλα αποτελούν μια από τις πιο υποσχόμενες και αποτελεσματικές τεχνολογικές μεθόδους για την ανίχνευση εχθρικού και ακατάλληλου λόγου σε στίχους τραγουδιών. Όμως, παρά τις σημαντικές προόδους στα νευρωνικά μοντέλα για την ανίχνευση εχθρικού λόγου, υπάρχουν σημαντικές προκλήσεις που πρέπει να ξεπεραστούν. Η συλλογή επαρκών ποιοτικών δεδομένων είναι δύσκολη, ιδιαίτερα για λιγότερο διαδεδομένες γλώσσες, και μπορεί να εγείρει ηθικά ζητήματα. Η ποιότητα των δεδομένων είναι κρίσιμη, καθώς κακή επισήμανση ή μεροληψία μπορεί να οδηγήσουν σε λανθασμένα αποτελέσματα. Επίσης, η πολυπλοκότητα και οι διαφορετικές μορφές εχθρικού λόγου καθιστούν δύσκολη τη δημιουργία ενός γενικού μοντέλου.

Τέλος, απαιτείται περισσότερη έρευνα για την ισορροπία μεταξύ ελευθερίας λόγου και αναγνώρισης εχθρικού περιεχομένου, ώστε να υποστηριχθεί η αποτελεσματική και δίκαιη εφαρμογή τέτοιων συστημάτων στη μουσική βιομηχανία, αλλά και, συνεχής βελτιώσεις στα δεδομένα και τα μοντέλα για ακριβέστερα αποτελέσματα.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Explicit Content Detection in Music Lyrics Using Machine Learning. 2018 IEEE International Conference on Big Data and Smart Computing.
- [2] Correlation Analysis Between Hate Speech and Audio Components of Songs Using Supervised Learning. Joonhyung Bae, Dongju Park.
- [3] C. Laurier, J. Grivolla and P. Herrera: "Multimodal Music Mood Classification Using Audio and Lyrics," In Proceedings of the International Conference on Machine Learning and Applications, 2008.
- [4] Pons, Jordi & Serra, Xavier. (2019). musicnn: Pretrained convolutional neural networks for music audio tagging.
- [5] Moon, Jihyung & Cho, Wonik & Lee, Junbum. (2020). BEEP! Korean Corpus of Online News Comments for Toxic Speech Detection.
- [6] F. Poletto, V. Basile, M. Sanguinetti, C. Bosco, and V. Patti, "Resources and benchmark corpora for hate speech detection: A systematic review," Language Resources and Evaluation, vol. 55, no. 2, p. 477–523, 2020.
- [7] Hu, X., & Downie, J. S. (2010). When lyrics outperform audio for music mood classification: A feature analysis. In ISMIR.
- [8] Malheiro, R., Panda, R., Gomes, P., & Paiva, R. P. (2016). Emotionally-relevant features for classification and regression of music lyrics. IEEE Transactions on Affective Computing.
- [9] Hu, X., Downie, J. S., & Ehmann, A. F. (2009). Lyric text mining in music mood classification. American Music.
- [10] Pohle, T., Pampalk, E., & Widmer, G. (2005). Evaluation of frequently used audio features for classification of music into perceptual categories. In Proceedings of the Fourth International Workshop on Adaptive Multimedia Retrieval.
- [11] Yang, D., & Lee, W. S. (2009). Music emotion identification from lyrics. In Proceedings of the IEEE International Symposium on Multimedia.
- [12] Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. ACM Computing Surveys (CSUR).

- [13] Chen, Y., Zhou, Y., Zhu, S., & Xu, H. "Detecting offensive language in social media to protect adolescent online safety." Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom). IEEE, 2012.
- [14] P. N. Juslin and P. Laukka: "Expression, Perception, and Induction of Musical Emotions: A Review and A Questionnaire Study of Everyday Listening," Journal of New Music Research, Vol. 33, No. 3, pp. 217-238, 2004.
- [15] X. Hu and J. S. Downie, "Improving mood classification in music digital libraries by combining lyrics and audio," Proc. Tenth Ann. joint conf. on Digital libraries, pp. 159-168, 2010.
- [16] Çano, E., & Morisio, M. (2017). Moodylyrics: A sentiment annotated lyrics dataset. ACM Transactions on Multimedia Computing, Communications, and Applications.
- [17] da Silva Malheiro, R. M. (2016). Emotion-based analysis and classification of music lyrics. Universidade de Coimbra.
- [18] M. M. Bradley and P. J. Lang. Affective norms for English words (ANEW): Stimuli, instruction manual, and affective ratings. Technical report, Center for Research in Psychophysiology, University of Florida, Gainesville, Florida, 1999.
- [19] Ballard, Mary E., and Steven Coates. "The immediate effects of homicidal, suicidal, and non-violent heavy metal and rap songs on the moods of college students." Youth & Society 27.2 (1995): 148-168.
- [20] Swati Agarwal and Ashish Sureka. 2015. Using KNN and SVM based one-class classifier for detecting online radicalization on Twitter. In Proceedings of the International Conference on Distributed Computing and Internet Technology. Springer, 431–442.
- [21] Singhi, A. (2015). Lyrics matter: Using lyrics to solve music information retrieval tasks. University of Waterloo.
- [22] F. Vignoli, "Digital Music Interaction concepts: a user study," Proc. of the 5th Int. Conference on Music Information Retrieval, 2004.
- [23] Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y "Abusive language detection in online user content." Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2016.
- [24] Arnett, Jeffrey. "Heavy metal music and reckless behavior among adolescents." Journal of youth and adolescence 20.6 (1991): 573-592.

[25] Rospocher, M. Detecting explicit lyrics: a case study in Italian music. *Lang Resources & Evaluation* 57, 849–867 (2023).

[26] Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Exploration Newsletter*, 6(1), 20–29.

[27] Bergelid, L. (2018). Classification of explicit music content using lyrics and music metadata. Master's thesis, KTH, School of Electrical Engineering and Computer Science (EECS).

[28] Bloehdorn, S., & Hotho, A. (2006). Boosting for text classification with semantic features. In B. Mobasher, O. Nasraoui, B. Liu, & B. Masand (Eds.), *Advances in web mining and web usage analysis* (pp. 149–166). Heidelberg: Springer.

[29] Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, and Vasudeva Varma. 2017. Deep learning for hate speech detection in tweets. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 759--760.

[30] Paula Fortuna and Sérgio Nunes. 2018. A Survey on Automatic Detection of Hate Speech in Text. *ACM Comput. Surv.* 51, 4, Article 85 (July 2018), 30 pages.

[31] M. Mandel, G. Poliner, and D. Ellis: "Support Vector Machine Active Learning for Music Retrieval," *Multimedia Systems*, Vol. 12, No. 1, pp. 3-13, 2006.

[32] X. Hu, J. S. Downie, C. Laurier, M. Bay, and A. Ehmann: "The 2007 MIREX Audio Music Classification Task: Lessons Learned," *Proceedings of the International Conference on Music Information Retrieval*, 2008.

[33] J.-J. Aucouturier and F. Pachet. Music similarity measures: What's the use? In Ircam, editor, *Proceedings of the 3rd International Symposium on Music Information Retrieval*, Paris, France, October 13-17 2002.

[34] Mary E. Ballard and Steven Coates, "The im-mediate effects of homicidal suicidal and non-violent heavy metal and rap songs on the moods of college students", *Youth & Society*, vol. 27, no. 2, pp. 148-168, 1995.

[35] Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.* 2, 160 (2021).

- [36] Swati Agarwal and Ashish Sureka. 2015. Using KNN and SVM based one-class classifier for detecting online radicalization on Twitter. In Proceedings of the International Conference on Distributed Computing and Internet Technology. Springer, 431–442.
- [37] Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, and Vasudeva Varma. 2017. Deep learning for hate speech detection in tweets. In Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, 759–760.
- [38] Vaglio, A., Hennequin, R., Moussallam, M., Richard, G., & d'Alche-Buc, F. (2020). Audio-Based Detection of Explicit Content in Music. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
- [39] Rospocher, M. (2020). Explicit song lyrics detection with subword-enriched word embeddings. Expert Systems with Applications, 113749.
- [40] Michael Fell, Elena Cabrio, Michele Corazza, Fabien Gandon. Comparing Automated Methods to Detect Explicit Content in Song Lyrics. RANLP 2019 - Recent Advances in Natural Language Processing, Sep 2019, Varna, Bulgaria.
- [41] Chin, H.; Kim, J.; Kim, Y.; Shin, J.; Yi, M.Y. Explicit Content Detection in Music Lyrics Using Machine Learning. In Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing, BigComp 2018, Shanghai, China, 15–17 January 2018; IEEE Computer Society: Washington, DC, USA, 2018; pp. 517–521.
- [42] P. Badjatiya, S. Gupta, M. Gupta, and V. Varma, “Deep Learning for Hate Speech Detection in Tweets”, in Proc. 26th Int. Conf. World Wide Web Companion - WWW 17 Companion, pp. 759–760, 2017.
- [43] R. Wang, Z. Li, J. Cao, T. Chen, and L. Wang, “Convolutional recurrent neural networks for text classification,” in 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–6, 2019.
- [44] P. K. Roy, A. K. Tripathy, T. K. Das, and X.-Z. Gao, “A framework for hate speech detection using deep convolutional neural network,” IEEE Access, vol. 8, p. 204951–204962, 2020.
- [45] D. C. Asogwa, C. I. Chukwuneke, C. C. Ngene, and G. N. Anigbogu, “Hate speech classification using svm and naive bayes,” Mar 2022.
- [46] A. G. D'Sa, I. Illina, and D. Fohr, “Bert and fasttext embeddings for automatic detection of toxic speech,” in 2020 International MultiConference on: “Organization of Knowledge and Advanced Technologies” (OCTA), pp. 1–5, 2020.

[47] T. Davidson, D. Warmley, M. Macy, and I. Weber, "Automated Hate Speech Detection and the Problem of Offensive Language," in Eleventh international aaai conference on web and social media, 2017.