

Προτεινόμενο Θέμα Διπλωματικής Εργασίας

Από

Ευστάθιος Σταματάτος
Καθ. Πανεπιστημίου Αιγαίου

Α. Επαλήθευση Ομιλητή από Μεταγραφές Ομιλίας (Speaker Verification from Speech Transcripts)

Περιγραφή:

Σε διάφορες εφαρμογές είναι επιθυμητό να εξακριβώσουμε την ταυτότητα ενός ομιλητή. Η περιοχή της επαλήθευσης ομιλητή (speaker verification) παραδοσιακά βασίζεται σε τεχνικές επεξεργασίας του ηχητικού σήματος. Σε περιπτώσεις που το ηχητικό σήμα παρουσιάζει θόρυβο ή κάποιος προσπαθεί να μιμηθεί την φωνή ενός ομιλητή, είναι πιο αποτελεσματικό να βασιστούμε στο κείμενο της απομαγνητοφώνησης (transcript) και να αναλυθεί το τι λέει κάποιος (περιεχόμενο) και κυρίως πώς το λέει (ύφος). Στα πλαίσια της εργασίας θα μελετηθούν τεχνικές αναπαράστασης κειμένου που εστιάζουν στο ύφος (που περιλαμβάνουν και παραδοσιακές τεχνικές ν-γραμμάτων και μεθόδους που βασίζονται σε βαθιά νευρωνικά δίκτυα και προ-εκπαιδευμένα γλωσσικά μοντέλα) και θα υλοποιηθεί προσέγγιση που θα συγκρίνει δύο κείμενα μεταγραφής ομιλίας (speech transcripts) για να αποφανθεί αν προέρχονται από τον ίδιο ή διαφορετικό ομιλητή. Ιδιαίτερη προσοχή θα δοθεί σε περιπτώσεις που το θέμα ενός ζεύγους κειμένων ταιριάζει ή δεν ταιριάζει για να φανεί κατά πόσο αυτό επηρεάζει την επίδοση του συστήματος επαλήθευσης ομιλητή. Θα διεξαχθεί αναλυτική πειραματική αξιολόγηση της προσέγγισης με βάση υπάρχουσες συλλογές δεδομένων για να εξαχθούν συμπεράσματα σχετικά με την αποτελεσματικότητα (effectiveness), ευρωστία (robustness), και υπολογιστικό κόστος (efficiency).

Στόχοι και παραδοτέα:

Η εργασία αποσκοπεί στην μελέτη των υπάρχοντων προσεγγίσεων αναπαράστασης κειμένου εστιάζοντας στο ύφος, την ανάπτυξη μεθόδου επαλήθευσης ομιλητή χρησιμοποιώντας τις μεταγραφές ομιλίας και την πειραματική αξιολόγηση της μεθόδου.

Παραδοτέα:

1. Πηγαίος κώδικας υλοποίησης
2. Αναλυτική αναφορά

Μαθησιακά αποτελέσματα:

Κατανόηση τεχνικών αναπαράστασης κειμένου εστιάζοντας στο ύφος.

Εξοικείωση με μεθόδους δυαδικής ταξινόμησης εστιάζοντας στην επαλήθευση ομιλητή/συγγραφέα.

Εξοικείωση με μεθόδους πειραματικής αξιολόγησης σε εργασίες δυαδικής ταξινόμησης.

References

1. Cristina Aggazzotti, Nicholas Andrews, and Elizabeth Allyn Smith. 2024. Can authorship attribution models distinguish speakers in speech transcripts? Transactions of the Association for Computational Linguistics, 12:875–891.
2. Yizhe Yang, Palakorn Achananuparp, Heyan Huang, Jing Jiang, and Ee-Peng Lim. 2024. Speaker Verification in Agent-generated Conversations. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, 5655–5676.
3. E.K. Sergidou, Rolf Ypma, Johan Rohdin, Marcel Worring, Zeno Geradts, and Wauter Bosma. 2024. Fusing linguistic and acoustic information for automated forensic speaker comparison. Science and Justice 64, 5 (2024), 485 – 497.