Κύριε Ψαράκη συγνώμη που δεν σας έστειλα την προηγούμενη βδομάδα.  
Η ιδέα μου είναι η εξής: Να χρησιμοποιήσω το state of the art δίκτυο το cosmic σαν βάση για το δικό μου δίκτυο. Συγκεκριμένα αυτό κάνει χρήση μόνο του modality του text και είναι αρκετά σύνθετη η αρχιτεκτονική του. Χρησιμοποιεί πρώτα ένα pretrained δίκτυο Roberta ( θα πρέπει να διαβάσω καλύτερα τι είναι αυτό) για context independent features του text και μετά χρησιμοποιεί το cosmic δίκτυο για να κατηγοριοποιήσει κάθε πρόταση στο συναίσθημα της. Το cosmic για κάθε πρόταση έχει 5 states:

To context state για την γενική πληροφορία , To internal state για το τι νιώθουν , Το intent state για το τι θέλουν και το External state για το τι δείχνουν πως συμπεριφέρονται ( εδώ παίζει ρόλο και το multimodality διότι το οπτικοακουστικό υλικό σε αυτό μας βοηθάει).

Εγώ σκέφτηκα να τροποποιήσω το δίκτυο αυτό προσθέτοντας audio features (για να εκμεταλλευτώ και τον ήχο όπως είπαμε) στο external state το οποίο διαπραγματεύεται τις εκφράσεις και τις αντιδράσεις των ομιλητών. Τα audio features που θα προσθέσω θα είναι βασισμένα πάνω στο πρώτο paper που μου στείλατε, δηλαδή ενός δικτύου που χρησιμοποιεί text και ήχο και θα είναι τα εξής:

A) Pitch γενική πληροφορία , B) Harmonics για θυμό και ενθουσιασμό, Γ) Speech energy (πόσο φωνάζει) για να ξεχωρίζει αρνητικά συναισθήματα όπως θυμό vs θλίψη, Δ) Pause τα κενά ανάμεσα στις λέξεις για να δούμε με πόση δυναμικότητα μιλάει, Ε) Central moments αυτό θα ήθελα να μου το εξηγήσετε.

A text on a page

AI-generated content may be incorrect.

Φαντάζομαι καλύτερα είναι να έρθω από εκεί να το συζητήσουμε, η αλήθεια είναι πως δεν ξέρω εάν η ιδέα αυτή είναι κατάλληλη, δεν ξέρω αν είναι πολύ εύκολη για διπλωματική ή πολύ σύνθετη ή ικανοποιητική.