Αναγνώριση Προτύπων

Θέμα: Αναγνώριση νοημάτων Ελληνικής Νοηματικής Γλώσσας

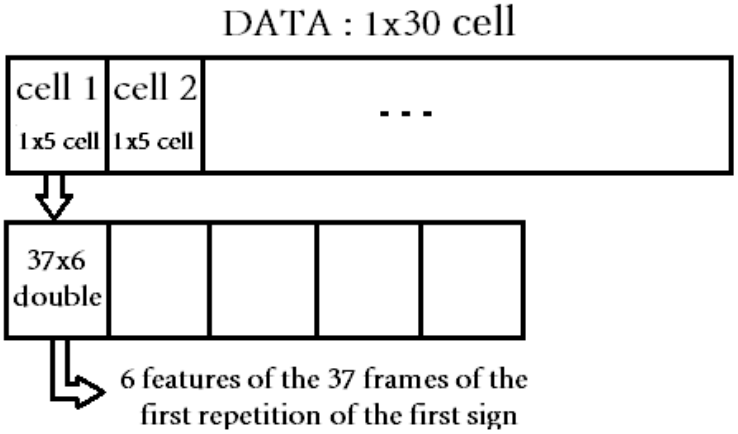
Εισαγωγή

Σ'αυτήν την εργασία θα προσπαθήσουμε να δημιουργήσουμε ένα αυτοματοποιημένο σύστημα αναγνώρισης 30 νοημάτων της ελληνικής νοηματικής γλώσσας. Για την δημιουργία ενός τέτοιου συστήματος θα χτησιμοποιήσουμε δεδομένα τα οποία παρέχονται απο το εργαστήριο του μαθήματος καθώς και θεωρία και τεχνικές αναγνώρισης προτύπων. Η εργασία αυτή θα περιέχει αρκετά μέρη οπου αρχικά θα χρησιμοποιήσουμε ένα απλό σύστημα που χρησιμοποιεί μόνο τα δοσμένα χαρακτηριστικά, ενώ στη συνέχεια θα εμπλουτίζουμε το σύστημά μας σταδιακά με νέα χαρακτηριστικά αλλα και ιδέες ωστε να πετύχουμε μεγαλύτερη γενίκευση και ακρίβεια κατηγοροποίησης και τελικά να δημιουργήσουμε ένα αρκετά ανεξάρτητο σύστημα που θα μπορεί να εκπαιδευτεί παρέχοντάς του μόνο τα βασικά χαρακτηριστικά των νοημάτων και να επιτυγχάνει καλή απόδοση σε ένα ευρύτερο σύνολο δεδομένων ελέγχου.

Μέρος 1 : Εξαγωγή δοσμένων χαρακτηριστικών, εκπαίδευση και έλεγχος

(Επιφανειακή μερική κάλυψη των E1 E5 και E6)

Σ'αυτό το πρώτο μέρος θα εξάγουμε απλά τα χαρακτηριστικά που δίνονται στο αρχείο FeatureTranscriptions.txt χωρίς να επαυξήσουμε τον χώρο εισόδου με νέα δεδομένα συνδυαζόμενα απο τα ήδη υπάρχοντα. Έτσι κάθε νόημα θα περιγράφεται απο ένα σύνολο διανυσμάτων ίσο με το πλήθος των πλαισίων που το χαρακτηρίζουν και κάθε διάνυσμα θα χαρακτηρίζεται απο 6 συνιστώσες οι οποίες περιγράφουν τη συντεταγμένη χ του κεφαλιού, την συντεταγμένη y του κεφαλιού, την συντεταγμένη x του δεξιού χεριού, την συντεταγμένη y του δεξιού χεριού, την συντεταγμένη x του αριστερού χεριού και την συντεταγμένη y του αριστερού χεριού. Η λογική την οποία θα ακολουθήσουμε στον τρόπο διαχείρισης των δεδομένων θα έιναι παρόμοια με τον τρόπο που ακολουθήθηκε στη 2η εργαστηριακή άσκηση. Συγκεκριμένα θα φροντίσουμε όλα τα δεδομένα , και των δύο ομιλητών, να είναι αποθηκευμένα sto 1x30 cell DATA.mat οπου κάθε κελί αντιστοιχεί σε ένα νόημα. Κάθε ένα απο αυτά τα κελιά περιέχει επιπλέον 5+ κελιά (όσες και οι επαναλήψεις του νοήματος) τα οποία με τη σειρά τους περιέχουν υπο τη μορφή πινάκων Tx6 τα χαρακτηριστικά διανύσματα των πλαισίων, οπου T το πλήθος των πλαισίων που αντιστοιχεί σε μια επανάληψη του συγκεκριμένου νοήματος. Για πάράδειγμα το πρώτο κελί του DATA.mat θα θέλουμε να περιέχει 5 επιπλέον κελιά (εφόσον δεν έχουμε επιπλέον signer για το νόημα ABROAD). Το πρώτο κελί θα είναι ένας πίνακας μεγέθους 37x6 οπου η κάθε γραμμή του πίνακα θα περιγράφει τα 6 βασικά χαρακτηριστικά. Το πλήθος γραμμών το πίνακα προκύπτει απο το δοσμένο InterSegmantation αρχείο το οποίο περιγράφει οτι η πρώτη επανάληψη του νοήματος ABROAD χαρακτηρίζεται απο τα πλαίσια 27:63 ,άρα σύνολο 37 πλαίσια. Φυσικά για να καταλήξουμε να έχουμε τη συγκεκριμένη δομή θα χρειαστεί αρκετή προεπεξεργασία. Στη συνέχεια ο σκοπός θα είναι να επαυξήσουμε τις 6 διαστάσεις των διανυσμάτων χαρακτηριστικών με επιπλέον χρήσιμα χαρακτηριστικά.



Σχήμα 1: Λογική αποθήκευσης χαρακτηριστικών

Η παραπάνω διαδικασια υλοποιείται απο τα m-files meros1a.m meros1b.m και meros1c.m. Σε πρώτη φάση θα χρησιμοποιήσουμε μόνο τα δεδομένα του βασικού ομιλητή και δεν θα λάβουμε υπ'όψιν τον δεύτερο ομιλητή ούτε κατα την εκπαίδευση αλλα ούτε κατα τον έλεγχο του συστήματος. Τα δεδομένα μόνο του πρώτου ομιλητή βρρίσκονται στο mat αρχείο DATA1.mat.

Για τη εκπαίδευση του συστήματος θα χρησιμοποιήσουμε ένα μέρος των δεδομένων ως training set και ένα μέρος ως testing set. Συγκεκριμένα επειδή σ'αυτήν την περίπτωση έχουμε μόνο 5 επαναλήψεις για κάθε νόημα, καλό θα είναι να χρησιμοποιήσουμε το 80% για δεδομένα εκπαίδευσης, δηλαδή τις 4 εκφωνήσεις, και το υπόλοιπο 20%, δηλαδή μια εκφώνηση απο το κάθε νόημα για έλεγχο. Βέβαια μπορούμε να κάνουμε δοκιμές και με άλλα ποσσοστά (όπως 60-40).

Για την εκπαίδευση θα χρησιμοποιήσουμε κρυφά μαρκοβιανά μοντέλα με τοπολογία left-right και απο 4 εως 8 καταστάσεις, αλλα και για τυχαίες τοπολογίες, ενώ θα γίνουν δοκιμές για μοντελοποίηση της κάθε κατάστασης με περισσότερες απο μια γκαουσιανές κατανομές (mixture of gaussians). Κάθε πείραμα θα το επαναλάβουμε 10 φορές ωστε να εξάγουμε μια μέση ακρίβεια ταξινόμησης. Όπως και στην 2η εργαστηριακή άσκηση θα παρατηρούμε ταυτόχρονα με την ακρίβεια και τον confusion matrix, τους όρους precision και recall όπου:

$$precision(i) = \frac{Σωστά\ ταξινομημένα\ νοήματα\ στην\ κατηγορία(i)}{Σύνολο\ ταξινομημένων\ νοημάτων\ σε\ αυτήν\ την\ κατηγορία(i)}$$

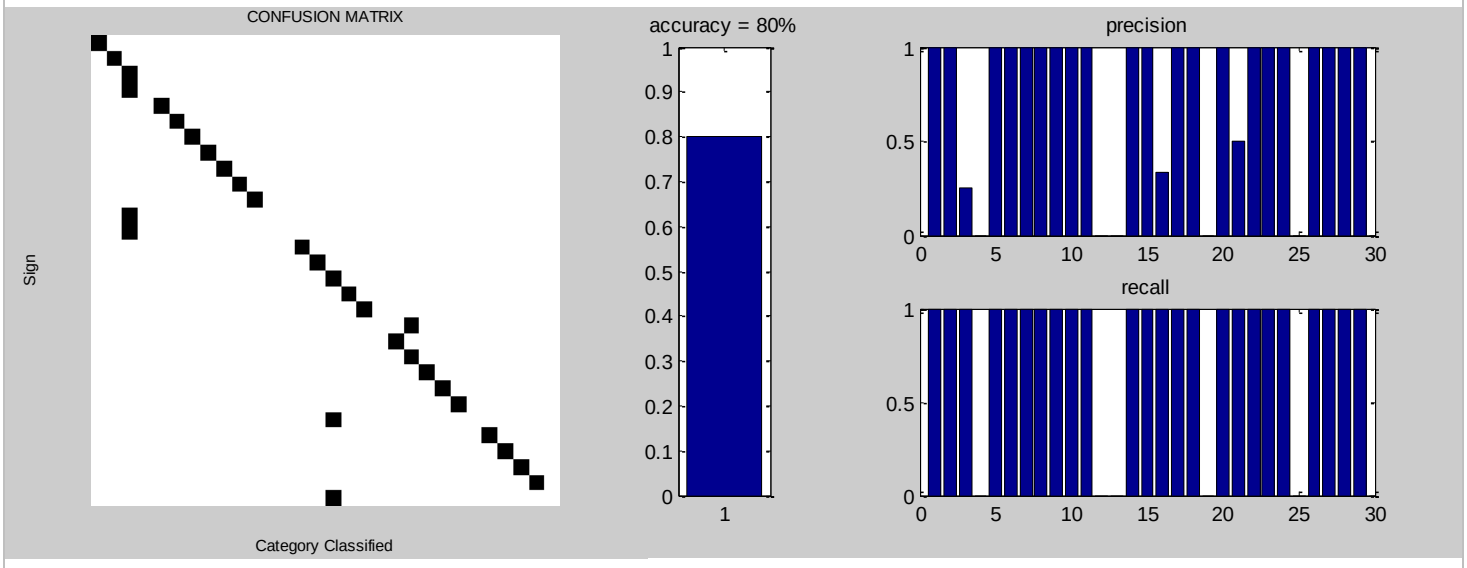
$$recall(i) = \frac{Σωστά\ ταξινομημένα\ νοήματα\ στην\ κατηγορία(i)}{Σύνολο\ επαναλήψεων\ του\ i\ νοήματος\ στο\ testing\ set}$$

Υπενθυμίζουμε τον τρόπο που παράγεται ο Confusion Matrix. Αντιστοιχούμε τα ¨νοήματα" σε μία σειριακή ακολουθία ακεραίων απο το "1" εως το πλήθος των νοημάτων. Για παράδειγμα το πρώτο νόημα αντιστοιχεί στο "1" το δεύτερο στο"2' και ούτω καθεξής. Η σειρά των νοημάτων είναι η σειρά με την οποία βρίσκονται στο αρχείο SignTranscriptions.txt . Για κάθε "νόημα" που αποτελεί δεδομένο του testing set γίνεται έλεγχος και ταξινομείται σε μία κατηγορία δηλαδή αναγνωρίζεται ώς κάποιο απ'αυτά. Αρχικά ο Confusion Matrix έχει όλα τα στοιχεία του μηδενικά. Τότε όταν το i-οστό νόημα αναγνωριστεί ως j-οστό , προσθέτουμε στο στοιχείο (i,j) του Confusion Matrix τη μονάδα. Έτσι μετά απο ένα πείραμα αναμένουμε ο Confusion Matrix να περιέχει μηδενικά και μη μηδενικά στοιχεία. Στην

ειδική περίπτωση που χωρίσουμε τα παραπάνω δεδομένα του πρώτου ομιλητή σε αναλογία 80%-20% training και test set οConfusion Matrix θα περιέχει μόνο άσσους και μηδενικά (γιατι ο ενας ομιλητής επαναλαμβάνει κάθε νόημα 5 φορές, αρα μόνο μια επανάληψη αντιστοιχεί σε test). Ωστόσο για να εξάγουμε πιο γενικευμένα συμπεράσματα σε κάθε περίπτωση θα επαναλάβουμε το πείραμα 10 φορές με τυχαίο διαχωρισμό των επαναλήψεων των νοημάτων σε training και test (κατα το ίδιο ποσσοστό). Τελικά μπορούμε να προσθέσουμε τους Confusion Matrices μεταξύ τους και στο τέλος να κανονικοποιήσουμε τις τιμές του απο 0 έως 1 ωστε να απεικονίσουμε τα αποτελέσματα ως μια grayscale εικόνα. Η Κανονικοποίηση αυτή γίνεται για κάθε γραμμή του πίνακα ξεχωριστά διαιρώντας τις τιμές των στοιχείων της γραμμής με το άθροισμα των τιμών των στοιχείων αυτών, που είναι το πλήθος των νοημάτων που ελέγχονται στην αντίστοιχη κατηγορία. Αυτό θα μας κάνει πιο εύκολη την παρακολούθηση και πιο άμεση την εξαγωγή συμπερασμάτων καθώς ιδανικά θα θέλαμε να επιτύχουμε μία είκόνα με λευκά τα διαγώνια στοιχεία και μάυρα όλα τα υπόλοιπα. Αυτό γιατι κάθε νόημα θα θέλαμε να αντιστοιχηθεί στην κατηγορία του οπότε μόνο τα διαγώνια στοιχεία των πινάκων θα έιχαν τιμή μονάδα. Προχωρώντας λίγο παραπέρα για λόγους πρακτικότητας αποφασίσαμε να απεικονίζουμε το συμπλήρωμα του πίνακα αυτού (CM = 1-CM) οπότε θα αναμένουμε ιδανικά να έχουμε μάυρα τα διαγώνια στοιχεία και λευκά τα μη διαγώνια.

Για τον υπολογισμό της ακρίβειας (accuracy) του μοντέλου προφανώς θα υπολογίσουμε απο τον Confusion Matrix στην αρχική του μορφή, το πηλίκο του αθροίσματος των διαγωνίων στοιχείων προς το άρθροισμα όλων των στοιχείων του πίνακα. Έπειτα για τον υπολογισμό του precision κάθε κατηγορίας λαμβάνουμε το πηλίκο της τιμής του στοιχείου που βρίσκεται πάνω στην γραμμή και στήλη του Confusion Matrix που αντιστοιχεί στην κατηγορία (νόημα) αυτή προς το άθροισμα των τιμών των στοιχείων που βρίσκονται πάνω στη στήλη αυτής της κατηγορίας. Στον υπολογισμό του recall το μόνο που αλλάζει είναι ο παρονομαστής οπου θα είναι προς το άθροισμα των τιμών των στοιχείων που βρίσκονται πάνω στη γραμμή αυτής της κατηγορίας

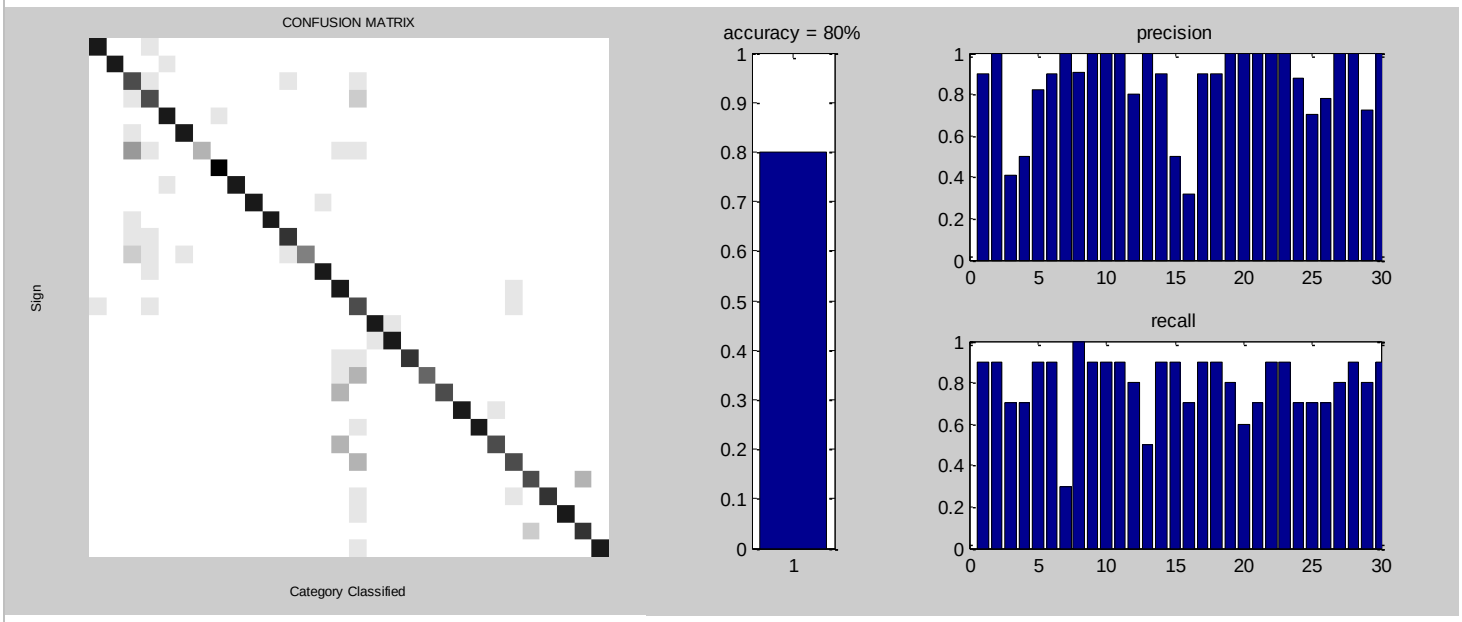
Αρχικά χρησιμοποιώντας τοπολογία μοντέλων left to right με πίνακα αρχικών πιθανοτήτων τέτοιων ώστε να επιτρέπει αρχική μετάβαση μόνο στην πρώτη κατάσταση (όπως και στη δεύτερη εργαστηριακή άσκηση), έγινε εκπαίδευση και έλεγχος με κρυφά μαρκοβιανά μοντέλα 5 καταστάσεων και μια γκαουσιανή ανα κατάσταση ενώ ο μέγιστος αριθμός επαναλήψεων του αλγορίθμου Expectation Maximization επιλέχθηκε να είναι 15.



Σχήμα 2: Αποτελέσματα για 5 καταστάσεις και 1 γκαουσιανή ανα κατάσταση για μια επανάληψη του πειράματος

Απο το παραπάνω σχήμα λοιπόν που εξήχθη απο μια επανάληψη του πειράματος μπορούμε να δούμε οτι το συνολικό ποσσοστό επιτυχίας είναι 80%. Βέβαια για μία μόνο επανάληψη δεν είναι ορθό να εξάγουμε ποσσοστά επιτυχίας (ειδικά τα precision και recall δεν έχουν και πολύ νόημα σε αυτήν την περίπτωση) ωστόσο μπορούμε να εξηγήσουμε λίγο αναλυτικότερα τα αποτελέσματα του Confusion Matrix βλέποντας για παράδειγμα οτι το 4ο "νόημα" που είναι το νόημα της λέξης AIRPORT να έχει ταξινομηθεί ώς το 3ο νόημα που αντιστοιχεί στη λέξη AIR. Επίσης ώς το νόημα της λέξης AIR αναγνωρίστηκαν και τα νοήματα 12 και 13 τα οποία αντιστοιχούν στις λέξεις ALREADY και ANGER αντίστοιχα. Γι'αυτό το λόγο βλέπουμε και το precision αυτής της κατηγορίας να είναι χαμηλό. Αντίστοιχα βλέπουμε οτι υπάρχουν άλλες τρείς λανθασμένες κατηγοριοποιήσεις.

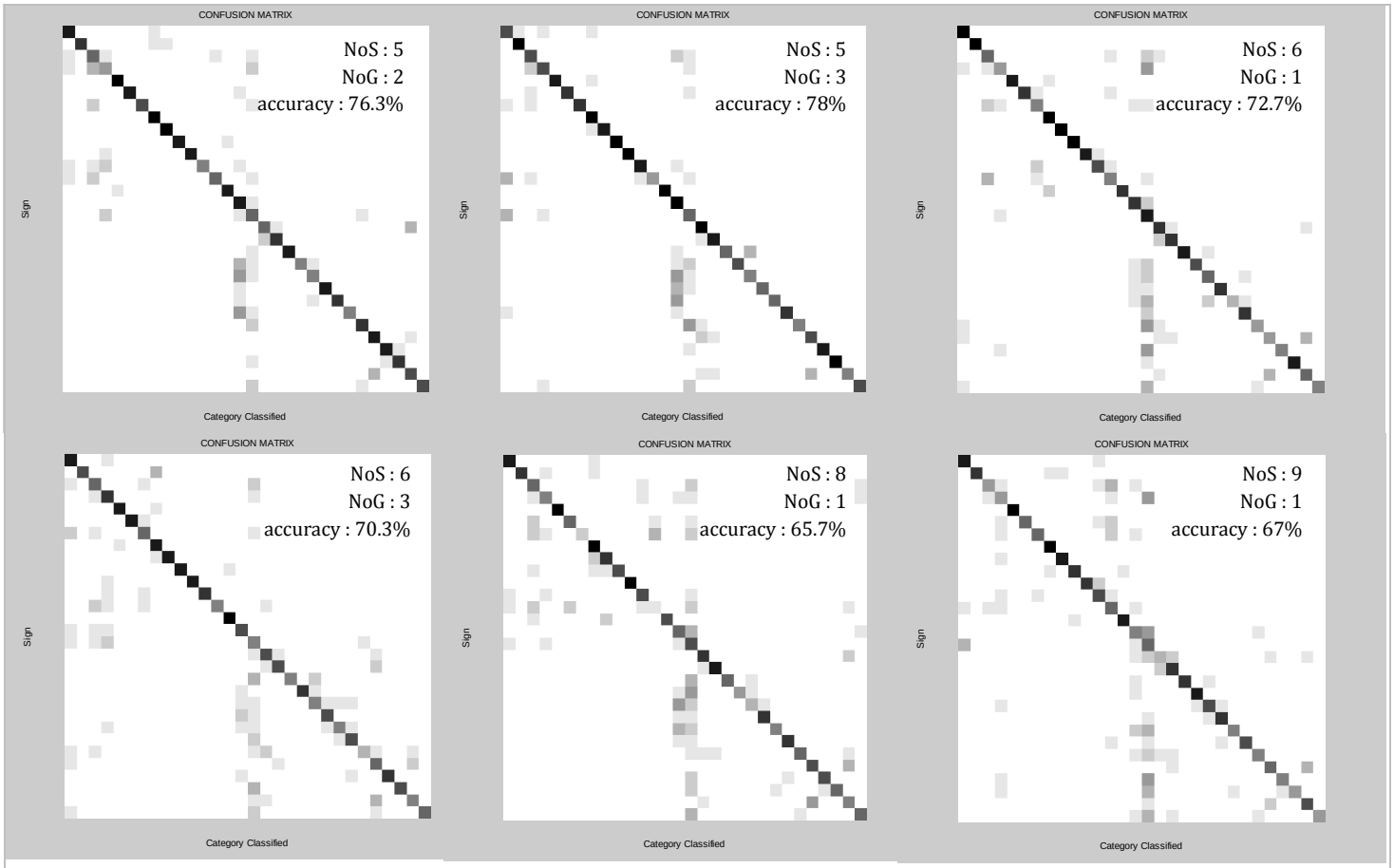
Στη συνέχεια το παραπάνω πείραμα επαναλήφθηκε 10 φορές με τυχαίο διαμοιρασμό των δεδομένων σε training και test set και λάβαμε έναν μέσο όρο ακρίβειας και παρουσιάζουμε τον Confusion Matrix σαν μια grayscale εικόνα με τρόπο που περιγράψαμε παραπάνω.



Σχήμα 3: Αποτελέσματα για 5 καταστάσεις και 1 γκαουσιανή ανα κατάσταση για 10 επαναλήψεις του πειράματος

Ας παρατηρήσουμε πρώτα τον confusion matrix του παραπάνω σχήματος. Αυτό που βλέπουμε αρχικά είναι οτι τα διαγώνια στοιχεία έχουν σίγουρα πιο σκούρα απόχρωση και αυτό σημαίνει οτι έχουμε σχετικά καλή ακρίβεια οπου πράγματι αυτό επιβεβαιώνεται και υπολογιστικά αφού όπως φαίνεται η ακρίβεια υπολογίστηκε να είναι 80%. Επίσης μπορούμε να δούμε οτι στις στήλες 3,4,15 και 16 έχουν ταξινομηθεί αρκετά λανθασμένα νοήματα και αυτό επιβεβαιώνεται απο το precision το οποίο είναι αρκετά χαμηλό και συγκεκριμένα κάτω απο 0.5 το οποίο πρακτικά σημαίνει οτι περισσότερα απο τα μισά νοήματα που κατηγοριοποιήθηκαν ως νοήματα αυτών των κατηγοριών ήταν εσφαλμένα. Τέλος βλέπουμε απο την 7η γραμμή του πίνακα πως τα νοήματα αυτής της κατηγορίας δεν κατηγοριοποιήθηκαν με αρκετή ακρίβεια καθώς το διαγώνιο στοιχείο δεν είναι το πιο σκούρο στοιχείο αυτής της γραμμής. Αυτό επιβεβαιώνεται απο το χαμηλό recall κάτω απο 0.5 που σημαίνει οτι περισσότερα απο τα μισά νοήματα αυτής της κατηγορίας ταξινομήθηκαν λανθασμένα σε άλλες κατηγορίες.

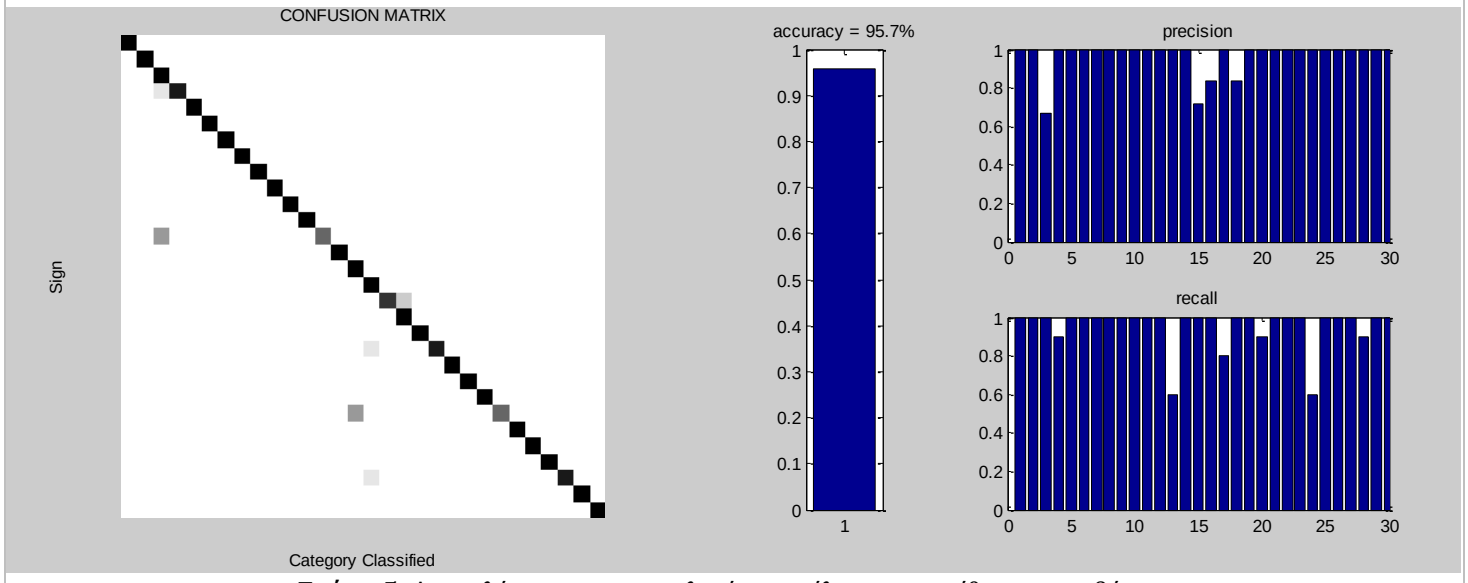
Πλέον αφού παρουσιάσαμε τον τρόπο αξιολόγησης των αποτελεσμάτων μπορούμε να εξάγουμε τα αποτελέσματα και για διαφορετικούς συνδυασμούς παραμέτρων. Τα αποτελέσματα αυτά για κάποιους ενδεικτικούς συνδιασμούς παρουσιάζονται στο σχήμα 4.



Σχήμα 4: Αποτελέσματα για διαφορους συνδυασμούς NoS , NoG για 10 επαναλήψεις του πειράματος και τοπολογία left to right

Απο τους παραπάνω πίνακες παρατηρούμε οτι αυξάνοντας το πλήθος των καταστάσεων ή και τις γακαουσιανές που περιγράφουν κάθε κατάσταση δεν αυξάνουμε την ακρίβεια της κατηγοριοποίησής μας, αντιθέτως η ακρίβεια μικραίνει ενώ η πολυπλοκότητα του μοντέλου αυξάνεται.

Στη συνέχεια ωφείλουμε να κάνουμε δοκιμές και με διαφορετική τοπολογία μοντέλων. Αυτή θα είναι να έχουμε ισοπίθανες μεταβάσεις απο κατάσταση σε κατάσταση αλλα ο πίνακας αρχικών πιθανοτήτων να επιτρέπει μετάβαση μόνο στην πρώτη κατάσταση έχοντας μονάδα στο πρώτο στοιχείο και μηδενικά όλα τα υπόλοιπα. Τα αποτελέσματα για μια γκουσιανή ανα κατάσταση και 5 καταστάσεις φαίνονται στο παρακάτω σχήμα.

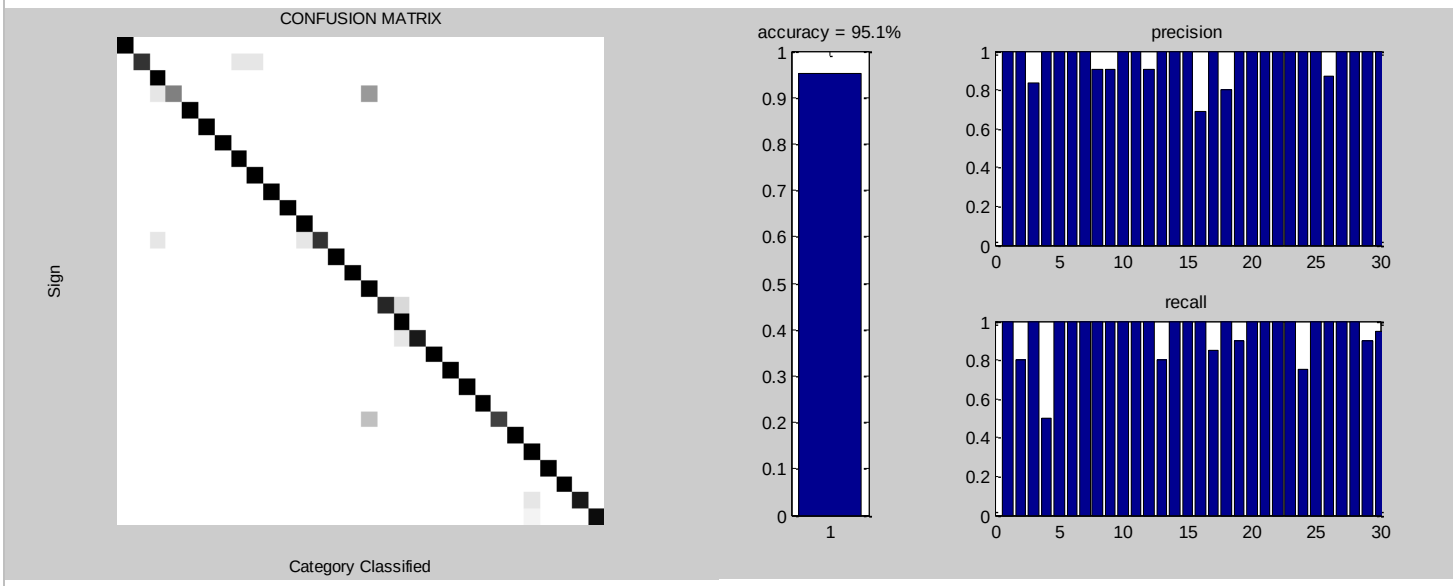


Σχήμα 5: Αποτελέσματα για τοπολογία μοντέλων με ισοπίθανες μεταβάσεις

Και πάλι για την εξαγωγή των παραπάνω αποτελεσμάτων επαναλάβαμε το πείραμα(10 φορές) με τυχαίο διαμοιρασμό του training και testing set καθε φορά. Απο αυτά παρατηρείται οτι η νέα τοπολογία είναι φανερά ισχυρότερη απο την προηγούμενη η οποία χρησιμοποιείται κυρίως για εκπαίδευση μοντέλων αναγνώρισης φωνής αλλα εδώ βλέπουμε πως δεν είναι η βέλτιστη τοπολογία.

Παρόλα αυτά τα αποτελέσματα των παραπάνω πειραμάτων δεν μπορούμε να θεωρήσουμε ακόμα οτί είναι αξιόπιστα καθώς περιέχουν δεδομένα αποκλειστικά απο έναν νοηματιστή. Γι 'αυτό το λόγο θα εισάγουμε και έναν δεύτερο νοηματιστή ο οποίος θα επαναλαμβάνει τα νοήματα απο το 16ο εως το 30ο πέντε φορές το καθένα. Έτσι θα δημιουργήσουμε ένα νέο cell αρχείο της ίδιας μορφής με το DATA1.mat το οποίο όμως απο το 16ο cell και μετά θα περιέχει 10 επαναλήψεις του αντίστοιχου νοήματος, 5 για κάθε νοηματιστή. Τελικά το νέο cell αποθηκεύεται με την ονομασία DATA_ALL.mat και η διαδικασία αυτή υλοποιείται μέσω του script concatenation_of_speakers.m. Επίσης στο αρχείο DATA_2nd_signer.mat αποθηκεύονται τα χαρακτηριστικά του δεύτερου νοηματιστή αποκλειστικά.

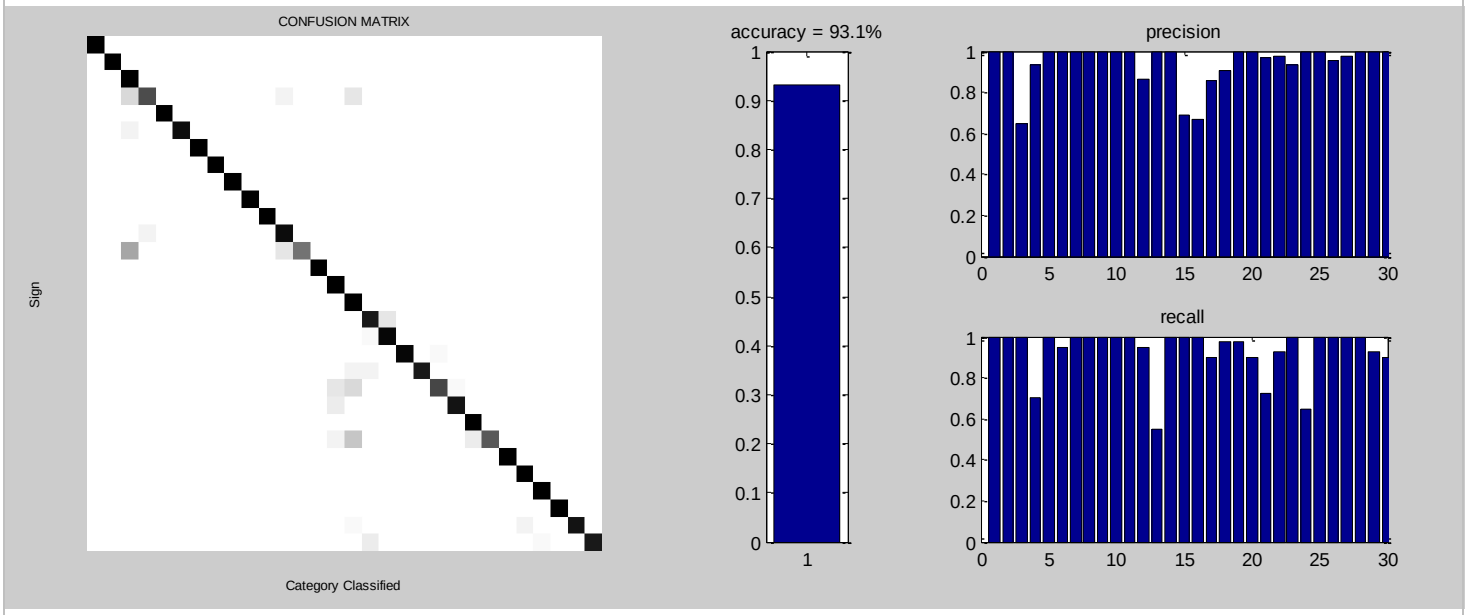
Στη συνέχεια επαναλάβαμε την παραπάνω διαδικασία με τα συνολικά δεδομένα με την νέα τοπολογία και με τις παραμέτρους που έχουμε επιλέξει ως βέλτιστες μεχρι αυτή τη στιγμή.



Σχήμα 7: Αποτελέσματα για εκπαίδευση και έλεγχο με τα δεδομένα και των δύο νοηματιστών

Απο το παραπάνω σχήμα βλέπουμε οτι το μοντέλο είναι ήδη αρκετά καλό αφού ακόμα και για τα νοήματα απο 16-30 επιτυγχάνουμε αρκετά καλές επιδόσεις. Να αναφέρουμε οτι το ποσσοστό επαναλήψεων των νοημάτων που χρησιμοποιούμε για εκπαίδευση του κάθε μοντέλου είναι το ίδιο ανεξάρτητα με το πλήθος των επαναλήψεων του κάθε νοήματων. Αυτό σημαίνει οτι ενω για τα νοήματα απο 1-15 με διαμοιρασμό δεδομένων σε 80-20% σε training και testing set θα έχουμε 4 επαναλήψεις για την εκπαίδευση και 1 για τον έλεγχο, στα νοήματα απο 16-30 που συνολικά θα έχουμε 10 επαναλήψεις για κάθε νόημα, οι 8 θα χρησιμοποιούνται για εκπαίδευση και οι υπόλοιπες δύο για έλεγχο. Στη συνέχεια ακούυθεί εκπαίδευση και έλεγχος του συστήματος με ποσσοστό διαμοιραμού των δεδομένων 60-40 για training και testing αντίστοιχα σε κάθε κατηγορία - μοντέλο και τ αποτελέσματα φαίνονται στο σχήμα 7.

Προφανώς όπως ήταν αναμενόμενο για μικρότερο σύνολο δεδομένων εκπαίδευσης, και αντίστοιχα μεγαλύτερο σύνολο δεδομένων ελέγχου έχουμε μικρότερες επιδόσεις.



Σχήμα 8: Αποτελέσματα για εκπαίδευση και έλεγχο με τα δεδομένα και των δύο νοηματιστών με διαμοιρασμό δεδομένων 60-40

Ωστόσο ακόμα και οι μεγάλες επιδόσεις που επιτυγχάνουμε μέσα απο αυτή τη διαδικασία και πάλι δεν είναι αξιόπιστες. Ερωτήματα όπως τι θα συμβεί στην περίπτωση που ο νοηματιστής δεν βρίσκεται στο κέντρο της εικόνας ή η εικόνα έχει διαφορετική ανάλυση αμέσως θα μας οδηγήσουν στο να αντιληφθούμε οτι οφείλουμε να γενικεύσουμε το σύστημά μας αναζητώντας νέα χαρακτηριστικά αυξάνοντας τη διάσταση του χώρου εισόδου και επεξεργάζοντας τα ήδη υπάρχοντα χαρακτηριστικά.

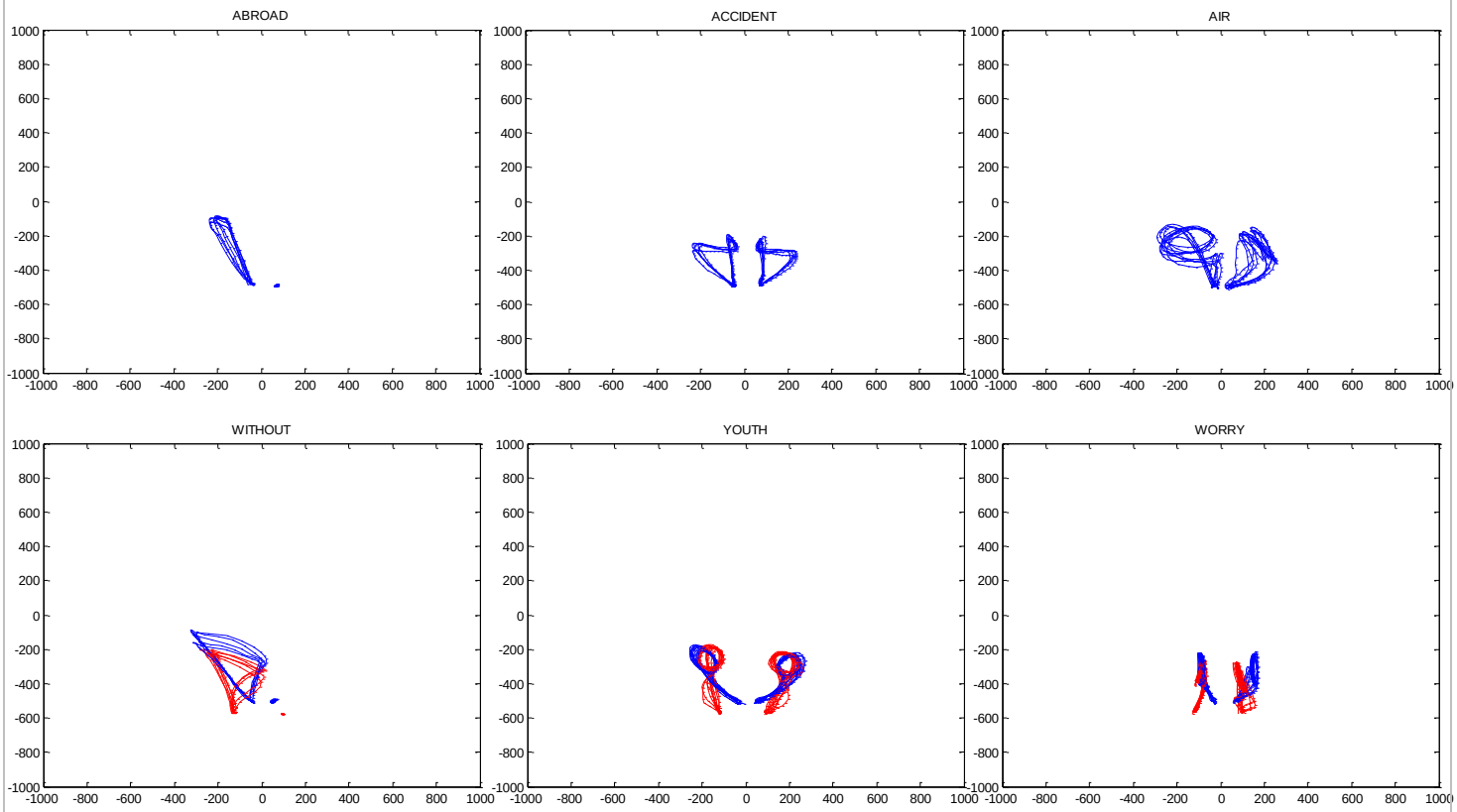
Μέρος 2 : Εξαγωγή νεων χαρακτηριστικών, τροποποίηση υπάρχοντων χαρακτηριστικών, εκπαίδευση και έλεγχος

(Πλήρης κάλυψη του E1, μερική κάλυψη των E3, E5 και E6)

Ξεκινάμε αυτό το μέρος προσπαθώντας να δώσουμε απαντήσεις στα όσα αναφέραμε στην τελευταία παράγραφο. Για να αποφύγουμε πιθανές αλλαγές και μεταβολές στον χώρο εισόδου που θα προέρχονται απο διαφορετικές αναλύσεις της εικόνας καθώς και μετακίνηση του νοηματιστή, μας οδηγεί αρχικά στην αναζήτηση ενός σημείου ισορροπίας για την περιγραφή των χαρακτηριστικών. Μια απλή σκέψη που θα εφαρμόσουμε είναι να αρχικά να συσχετίσουμε τις θέσεις των χεριών με τη θέση που βρίσκεται το κεφάλι. Συγκεκριμένα ώς νέα "σχετική" θέση του δεξιού χεριού θα θεωρούμε τη διαφορά της θέσης (x_δ,y_δ) όπως αυτή αναφέρεται στα αρχεία FeatureTranscritions με την θέση (x_κ,y_κ) του κεφαλιού. Αντίστοιχα και για το αριστερό χέρι. Έτσι θα έχουμε μια ανεξάρτητη περιγραφή των χαρακτηριστικών απο τη θέση του νοηματιστή στην εικόνα ενώ ταυτόχρονα δεν χάνουμε την πληροφορία των διαστάσεων που περιγράφουν τη θέση του κεφαλιού καθώς στην πραγματικότητα οι 4 νέες διαστάσεις είναι γραμμικά συσχετιζόμενες με αυτή. Συγκεκριμένα τα 4 νέα χαρακτηριστικά θα είναι $(x'_\delta , y'_\delta , x'_\alpha , y'_\alpha) = (x_\delta-x_\kappa , y_\delta-y_\kappa , x_\alpha-x_\kappa , y_\alpha-y_\kappa)$ οπου x'_δ η σχετική x συντεταγμένη του δεξιού χεριού και αντίστοιχα είναι αντιληπτοί και οι άλλοι συμβολισμοί. Στη συνέχεια κάτι άλλο που θα πρέπει να κάνουμε ωστε να γενικεύσουμε ακόμα περισσότερο τα χαρακτηριστικά είναι να διαιρέσουμε κάθε x' χαρακτηριστικό με μια παράμετρο σχετική με το μέγεθος x σε pixels της εικόνας (1440) απο την οποία προέρχονται τα χαρακτηριστικά αυτά ανώ αντίστοιχα τα y' χαρακτηριστικά με το μέγεθος y σε pixels της εικόνας (1080). Στη συγκεκριμένη περίπτωση θα μπορούσαμε να διαιρέσουμε με 1.44 και 1.08 αντίστοιχα. Αυτό θα μας οδηγήσει στο να έχουμε όλα τα

χαρακτηριστικά κανονικοποιημένα απο -1000 έως 1000. Κάτι τέτοιο είναι προφανές καθώς αποκλείεται για παράδειγμα η διαφορά x_{δ}-x_{κ} να είναι μεγαλύτερη κατα απόλυτο τιμή απο το πλήθος των εικονοστοιχείων που περιγράφουν αυτή τη διάσταση της εικόνας (απο απλή παρατήρηση ωστόσο των τιμών υπάρχει η πεποίθηση-βεβαιότητα οτι τα frames υποδειγματοληπτήθηκαν πρίν την εξαγωγή χαρακτηριστικών, μειώνοντας τη διάσταση της εικόνας στο μισό άρα οι συντελεστές αυτοί θα είναι 0.72 και 0.54 αντιστοιχα). Η παραπάνω παραμετροποίηση των χαρακτηριστικών υλοποιείται απο το script generalise_the_data.m και τα γενικευμένα δεδομένα αποθηκεύονται στο αρχείο DATA1_G.mat για τον πρώτο νοηματιστή, στο DATA_2nd_signer_G.mat για τον δεύτερο και στο DATA_ALL_G.mat για το σύνολο των δεδομένων.

Στη συνέχεια στο σχήμα 8 θα απεικονίσουμε κάποια απο τα χαρακτηριστικά και τον τρόπο που εξελίσσονται στο επίπεδο (x',y') οπου σε κάθε πλάισιο θα απεικονίζουμε όλες τις επαναλήψεις, με μπλέ χρώμα του 1ου ομιλητή και με κόκκινο χρώμα του δεύτερου.

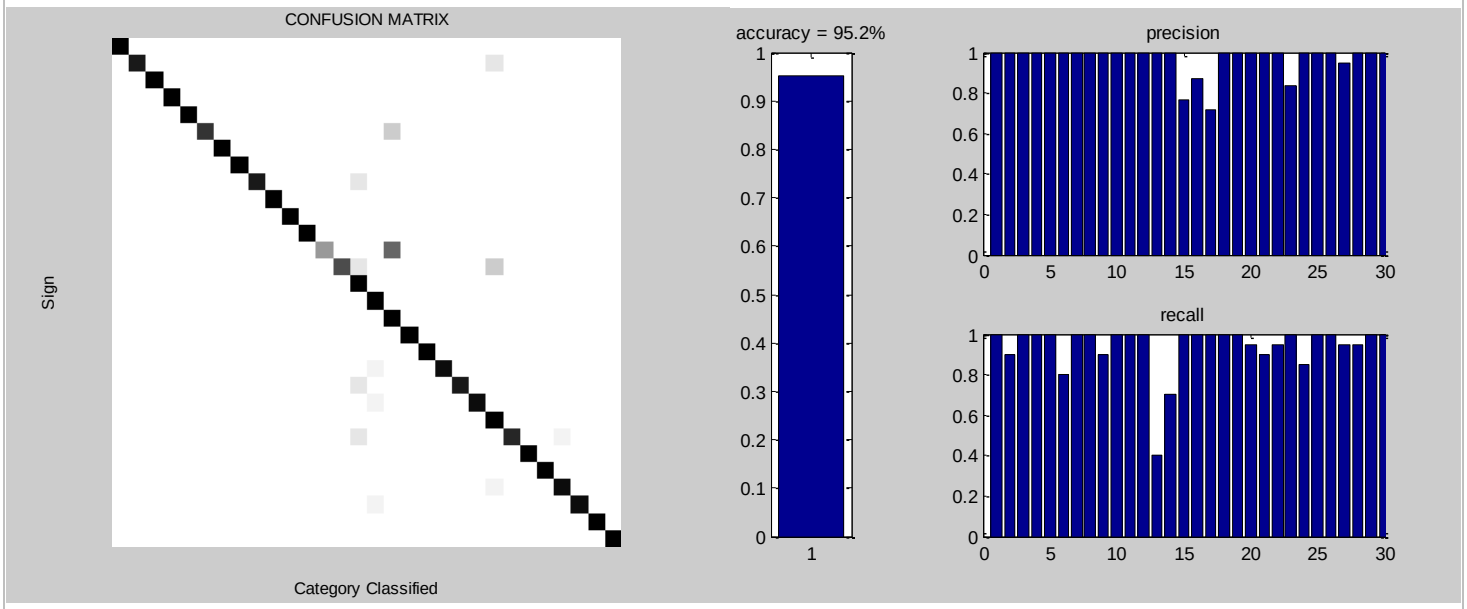


Σχήμα 9 : Απεικονίσεις των σχετικών κανονικοποιημένων θέσεων των χεριών

Στο παραπάνω σχήμα αυτό που παρατηρούμε είναι στην πρώτη γραμμή επαναλήψεις νοημάτων που είχαμε μόνο τον πρώτο νοηματιστή ενώ στην δεύτερη γραμμή είχαμε και δεδομένα απο τον δεύτερο νοηματιστή (κόκκινο χρώμα). Στην πρώτη στήλη στα νοήματα δεν χρησιμοποιείται το αριστερό χέρι, στη δεύτερη η κίνηση του αριστερού χεριού είναι συμμετρική και στην τρίτη αντισυμμετρική. Αυτό προς το παρόν το γνωρίζουμε παρατηρώντας τα αντίστοιχα βίντεο των νοημάτων και όχι απο κάποιον αυτοματοποιημένο τρόπο.

Είναι προφανές οτι τα δεδομένα του δεύτερου νοηματιστή διαφέρουν μερικώς απο του πρώτου αλλα ωστόσο υπάρχουν πολλές ομοιότητες ειδικά αν λάβουμε υπ'όψιν το γεγονός οτι πλέον ο τρόπος με τον οποίο εξάγουμε τα βασικά αυτά χαρακτηριστικά είναι αρκετά καλύτερος ωστε να μπορεί το σύστημα να εφαρμοστεί πάνω και σε άλλους νοηματιστές. Θα μπορούσαμε να συνεχίσουμε αυτήν την

παραμετροποίηση γενικέυοντας ακόμα περισσότερο τη μέθοδο αυτή αλλα θα αρκεστούμε σε αυτό το σημείο ωστε να παρατηρήσουμε τα πρώτα αποτελέσματα. Αυτά φαίνονται στο παρακάτω σχήμα οπου χρησιμοποιήσαμε διαχωρισμό δεδομένων 80-20 και τις βέλτιστες παραμέτρους μέχρι στιγμής.



Σχήμα 10 : Αποτελέσματα για κανονικοποιημένα δεδομένα

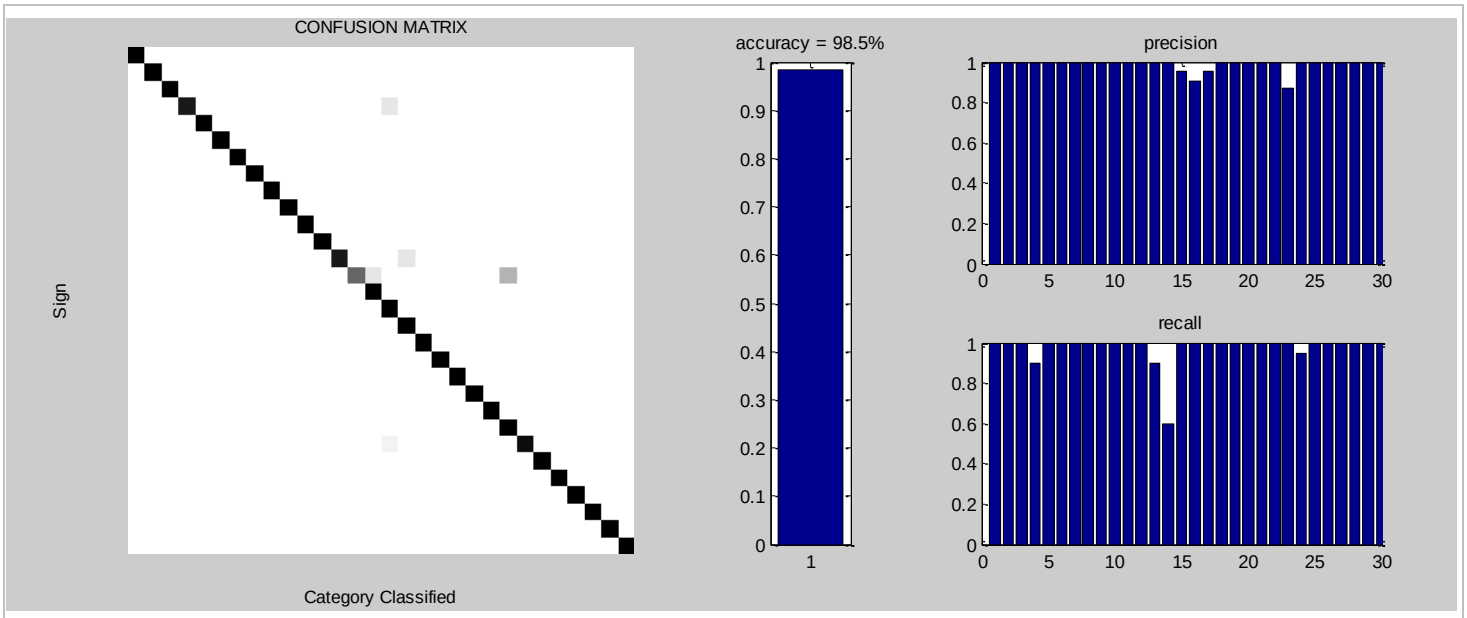
Παρατηρούμε λοιπόν οτι παρόλο που μειώσαμε τη διάσταση του χώρου εισόδου κατα δύο παρ'όλα αυτά τα αποτελέσματα είναι και πάλι ικανοποιητικά. Επίσης πλέον το μοντέλο μπορεί να δώσει καλά αποτελέσματα και για διαφορετικές βιντεοσκοπήσεις και έπειτα εξαγωγή των χαρακτηριστικών των νοημάτων καθώς δεν το αποτέλεσμα δεν εξαρτάται απο τη θέση του νοηματιστή στην εικόνα.

Στη συνέχεια θα εξάγουμε νέα χαρακτηριστικά απο τα ήδη υπάρχοντα ωστε να αυξήσουμε τη διάσταση του χώρου εισόδου ελπίζοντας οτι η νεα περιγραφή των πλαισίων θα μας οδηγήσει στην εκαπαίδευση ενος πλήρους συστήματος. Τα νεα χαρακτηριστικά που επιπλέγουμε θα είναι οι διαφορές των συντεταγμένων των χεριών, και οι ταχύτητές τους. Συγκέκριμένα τα χαρακτηριστικά φαίνονται στον παρακάτω πίνακα. Αυτό γιατί γνωρίζουμε πρακτικά τη σημαντικότητα της σχετικής θέσης των χεριών καθώς και των ταχυτήτων των χεριών (οι οποίες υπολογίστηκαν με χρήση της συνάρτησης diff) κατα την κίνησή τους απο τον νοηματιστή,

x'_{δ}	y'_{δ}	x'_{α}	y'_{α}	$x'_{\delta} - x'_{\alpha}$	$y'_{\delta} - y'_{\alpha}$	vx'_{δ}	vy'_{δ}	vx'_{α}	vy'_{α}

Πίνακας 1 : Διάνυσμα χαρακτηριστικών

Στη συνέχεια έγινε έλεγχος με το νέο αυτό σύνολο δεδομένων και το πείραμα επαναλήφθηκε 10 φορές. Τα αποτελέσματα φαίνονται στο σχήμα 11. Μπορούμε να δούμε λοιπόν οτι πραγματικά επιτύχαμε μια πολύ καλή ακρίβεια κατηγοριοποίησης του μοντέλου (98.5%) και καλό percision και recall σε όλες τις κατηγορίες εκτός απο την 14η κατηγορία που αντιστοιχεί στο νόημα ANSWER οπου το recall είναι αρκετά χαμηλό (0.6).



Σχήμα 11 : Αποτελέσματα για κανονικοποιημένα δεδομένα και προσαυξημένο διανυσμα χαρακτηριστικών

Με την αύξηση της διάστασης του διανύσματος χαρακτηριστικών με νέα τα οποία έχουν σημαντικό ρόλο όπως αντιλαμβανόμαστε κατα την κίνηση του νοηματιστή καταφέραμε να αυξήσουμε την ακρίβεια της ταξινόμησης αισθητά. Στη συνέχεια προσθέσαμε κι άλλα χαρακτηριστικά όπως σχετικές ταχύτητες χεριών , επιτάχυνση, μέτρο ταχυτήτων, και επαναλάβαμε το πείραμα ,όμως τα αποτελέσματα ακρίβειας μειώθηκαν αισθητά. Τελικά θα αρκεστούμε προς το παρόν σε αυτά τα 10 χαρακτηριστικά που περιγράφουν ικανοποιητικά το χώρο εισόδου οπου λόγω της αρχικής επεξεργασίας επιτύχαμε αρκετά καλή γενίκευση ενω προσθέτωντας και νέα χαρακτηριστικά έχουμε ήδη επιτύχει να εκπαιδεύσουμε ενα αρκετά δυνατό σύστημα αναγνώρισης των 30 αυτών νοημάτων.

Μέρος 3 : Ευρεση ενεργών περιοχών νοηματισμού

(Πλήρης κάλυψη του Ε2)

Ξεκινάμε αυτό το μέρος με το εξής σκεπτικό. Ψάχνουμε ένα κριτήριο με το οποίο θα αποφασίζουμε αν η περιοχή νοηματισμού είναι σημαντική ή όχι. Θεωρούμε οτι έχουμε στα χέρια μας μόνο τα αρχεία FeatureTranscriptions και έχουμε εξάγει τα βασικά χαρακτηριστικά κάθε αρχείου έχοντας έτσι ένα διάνυσμα 6 διαστάσεων για κάθε πλαίσιο.

Για να απαντήσουμε σε αυτό το ερώτημα υποθέτουμε αρχικά οτι ο νοηματιστής είναι ακίνητος και στη θέση ηρεμίας. Έτσι μπορούμε να θεωρήσουμε τις αρχικές συντεταγμένες ως συντεταγμένες ηρεμίας. Επίσης στη συνέχεια μπορούμε να υπολογίσουμε την ταχύτητα κατα x και κατα y και να επαυξήσουμε το διάνυσμα χαρακτηριστικών. Το δεξί χέρι είναι το κύριο χέρι και γι'αυτό δεν χρειάζεται να υπολογίσουμε τις ταχύτητες του αριστερού χεριού ή του κεφαλιού. Τελικά μπορούμε να κάνουμε κάποιους υπολογισμούς. Συγκεκριμένα υπολογίζουμε: την απόλυτη διαφορά μεταξύ συντεταγμένης x κεφαλιού απο τη συντεταγμένη αναφοράς x κεφαλιού, την απόλυτη διαφορά μεταξύ της συντεταγμένης y κεφαλιού απο τη συντεταγμένη αναφοράς y του κεφαλιού, τις αντίστοιχες απόλυτες διαφορές για το δεξί και αριστερό χέρι, την απόλυτη τιμή ταχύτητας κατα x του δεξιού χεριού και την απόλυτη τιμή ταχύτητας κατα y του δεξιού χεριού. Τελικά αθροίζουμε όλες αυτές τις τιμές και έχουμε ένα πολύ απλό αλλα αποτελεσματικό κριτήριο διαχωρισμού των πλαισίων απο ενεργά και μη ενεργά. Συγκεκριμένα :

$$J = |x_k - x_{kc}| + |y_k - y_{kc}| + |x_\delta - x_{\delta c}| + |y_\delta - y_{\delta c}| + |x_a - x_{ac}| + |y_a - y_{ac}| + |v_{x\delta}| + |v_{y\delta}|$$

όπου οι συμβολισμοί είναι οι προφανείς. Έτσι για κάθε πλαίσιο μπορούμε να βρούμε την τιμή του κριτηρίου αυτού και τελικά για κάθε βίντεο θα έχουμε έναν πίνακα στήλη ο οποίος σε κάθε γραμμή θα περιέχει την τιμή του κριτηρίου αυτού για το αντίστοιχο πλαίσιο.

Στη συνέχεια θα πρέπει να βρούμε ένα κατώφλι πάνω απο το οποίο θα θεωρούμε οτι το πλάισιο ανήκει σε ενεργή περιοχή. Στη συγκεκριμένη περίπτωση απο απλή επισκόπιση του πίνακα στήλης J για κάθε πλαίσιο παρατηρήθηκε οτι μια σχετικά καλή τιμή θα είναι η τιμή 80. Απο την παρατήρηση αυτή επιβεβαιώθηκε το γεγονός οτι η τιμή του κριτηρίου είναι αρκετά μικρή όταν ο νοηματιστής είναι ακίνητος και κοντά στη θέση ηρεμίας καθώς όλες οι απόλυτες τιμές είναι σχετικά μικρές. Βέβαια δεν θα είναι μηδενικές εκτός απο το πρώτο πλαίσιο και αυτό γιατι δεν είναι λογικό οτι ο νοηματιστής να επιστρέφει ακριβώς στο ίδιο σημείο απ'όπου ξεκίνησε (οπου ακόμα και να το έκανε δεν θα εξήγαμε πάντα τα ίδια ακριβώς σημεία απο τις τεχνικές όρασης υπολογιστών) με ακρίβεια εικονοστοιχείου. Επίσης είναι πού πιθανό να κινείται έστω και ελάχιστα π.χ 9 ή 10 εικονοστοιχεία ανα πλαίσιο(βέβαια αυτό εξαρτάται απο τις διαστάσεις τις εικόνας και την ανάλυσή της σε pixels). Συνολικά αν δώσουμε μια μέση ανοχή σε κάθε κατηγορία περίπου 10 εικονοστοιχεία καταλήγουμε για τις 8 κατηγορίες που αναφέραμε στο κριτήριό μας να θεωρούμε ώς κατώφλι την τιμή 80.

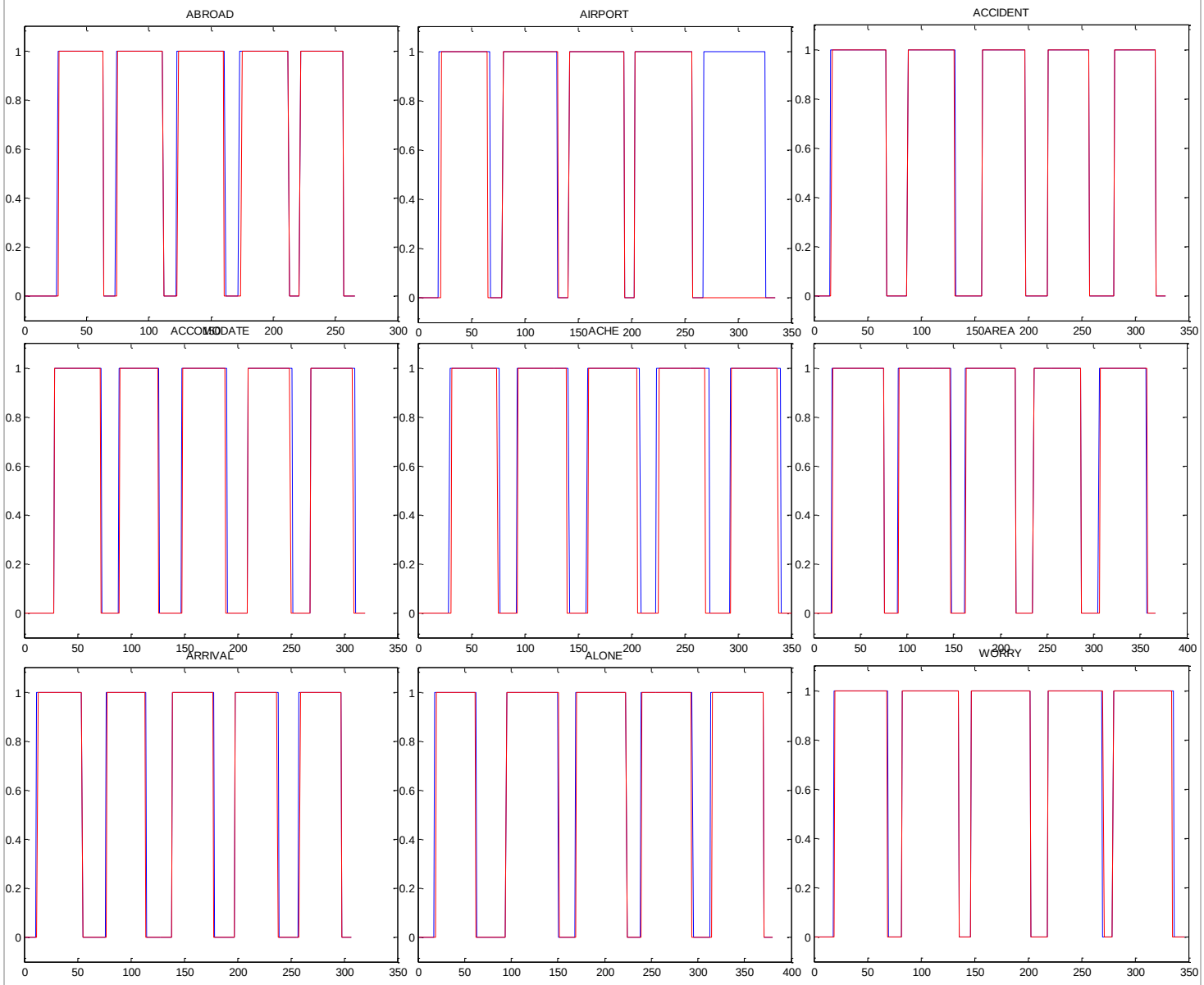
Στη συνέχεια με βάση αυτό το κατώφλι μπορούμε να κατασκευάσουμε έναν πίνακα στήλη ίδιων διαστάσεων που να έχει το ίδιο μέγεθος με τον J (και προφανώς με το πλήθος των πλαισίων για κάθε ομάδα νοημάτων-βίντεο) ο οποίος να έχει μηδενικά όπου ο J είναι μικρότερος ή ίσος με 80 και μονάδα όταν είναι μεγαλύτερος. Προφανώς οι άσσοι θα χαρακτηρίζουν τις ενεργές περιοχές και τα μηδενικά τις μη ενεργές. Εδώ θα σκεφτούμε ωστόσο να δώσουμε και ένα σχετικό περιθώριο λαμβάνοντας ώς ενεργά πλαίσια και τρια επιπλέον πλαίσια που είναι στα όρια των αλλαγών ωστε να ¨κρατήσουμε ¨ το κάθε νόημα απο την αρχή με κάποιο περιθώριο. Αυτό μπορούμε να το κάνουμε με χρήση της συνάρτησης imdilate του Matlab για 1 διάσταση ουσιαστικά αυξάνοντας τις περιοχές των άσσων κατα τρια στοιχεία πάνω και κάτω (βλέπε κώδικα) . Έστω οτι ονομάζουμε τον πίνακα αυτόν Ε.

Για να βρούμε στη συνέχεια τα πλαίσια στα οποία αρχίζει και τελειώνει ένα νόημα θα κάνουμε το εξής. Με χρήση των συναρτήσεων abs και diff υπολογίζουμε την απόλυτη τιμή της παραγώγου του πίνακα Ε (έστω πίνακας DE) οπου θα έχουμε παντού μηδενικά εκτός απο τις γραμμές εκείνες (δηλαδή τα πλάισια αφού κάθε γραμμή αντιστοιχεί σε ένα πλαίσιο) οπου είχαμε αλλαγή απο 0 σε 1 ή αντίστροφα. Βέβαια λόγω της διακριτότητας θα έχουμε διαφορά +- μια γραμμή ενδιαφέροντος. Τελικά με χρήση της συνάρτησης find πάνω στον DE θα έχουμε τα πλαίσια στα οποία αλλάξουμε απο ανενεργή περιοχή σε ενεργή και απο ενεργή σε ανενεργή. Λαμβάνοντας τα περιττά στοιχεία του αποτελέσματος της find θα έχουμε προφανώς τα πιθανά σημεία στα οποία ξεκινάει ένα νόημα και τα άρτια στοιχεία θα είναι τα στοιχεία στα οποία σταματάει το νόημα αυτό.

Όλη η παραπάνω διαδικασία υλοποιείται απο το script segmentation.m όπου τελικά τα αποτελέσματα αποθηκέυονται με τον ίδιο ακριβώς τρόπο με τον οποίο αποθηκεύονται όταν εξάγονται και απο τα αρχεία InterSegmentation στο script meros1b.m. Επίσης η διαδικασία αυτή πραγματοποιήθηκε μόνο για τον πρώτο ομιλητή και τελικά συγκρίνουμε τα εξαγόμενα πλαίσια ενδιαφέροντος που δίνονται με αυτά που εξάγουμε εμείς. Η σύγκριση αυτή για κάποια απο τα νοήματα φαίνεται στο σχήμα 12.

Σημειώνουμε οτι ενώ η μέθοδος δείχνει να λειτουργεί με πολύ καλό βαθμό προσέγγισης για σχεδόν όλα τα βίντεο παρόλα αυτά για 3 απ'αυτά δείχνει αδυναμία. Για παράδειγμα στην τέταρτη κατηγορία που αντιστοιχεί στο νόημα AIRPORT τα χέρια του νοηματιστή είναι αρκετά κοντά στο σημείο αναφοράς προς το τέλος του νοήματος και επίσης η ταχύτητα είναι πολύ μικρή. Το παραπανω δεν αποτελεί σημαντικό πρόβλημα κατα την εξαγωγή των περιοχών ενδιαφέροντος γιατί απλά τρείς απο

τις συνολικά 150 επαναλήψεις δε αναγνωρίζονται αλλα για τα συγκεκριμένα νοήματα αναγνωρίζεται τουλάχιστον μια ορθώς οριοθετημένη επανάληψη. τα αποτελέσματα φαίνονται στα παρακάτω διαγράμματα όπου απεικονίζουμε με 1 τις περιοχές ενδιαφέροντος και με 0 τις μη ενεργές περιοχές. Επίσης με μπλέ χρώμα φάινονται οι ενεργές περιοχές όπως έχουν δωθεί απο το εργαστήριο ενώ με κόκκινο είναι οι υπολογιζόμενες απο το δικό μας αυτοματοποιημένο σύστημα.



Σχήμα 12 : Εμφάνιση ενεργών, με τιμή "1" , και μη ενεργών περιοχών , με τιμή "0", απο τα δοσμένα δεδομένα , με μπλέ χρώμα, και απο τον αυτόματο τρόπο, με κόκκινο χρώμα για τυχαία νοήματα.

Απο τα παραπάνω διαγράμματα επιβεβαιώνουμε την ικανότητα του αλγορίθμου να βρίσκει ικανοποιητικά τις περιοχές ενδιαφέροντος. Οι μοναδικές περιπτώσεις που δεν βρέθηκαν όλες οι επαναλήψεις είναι για τα νοήματα AIRPORT , RUIN και MOVERIGHT. Επίσης τα πλαίσια έχουν μια πολύ μικρή απόκλιση απο1 εως 3 frames ως προς τα όρια των περιοχών κάτι που είναι αποδεκτό καθώς εξάλλου το πότε πραγματικά ξεκινάει να έχει ενδιαφέρον η κίνηση είναι ενα πρόβλημα που κείται στην υποκειμενική κρίση του παρατηρητή.

Μέρος 4 : Ενδονοηματική κατάτμηση

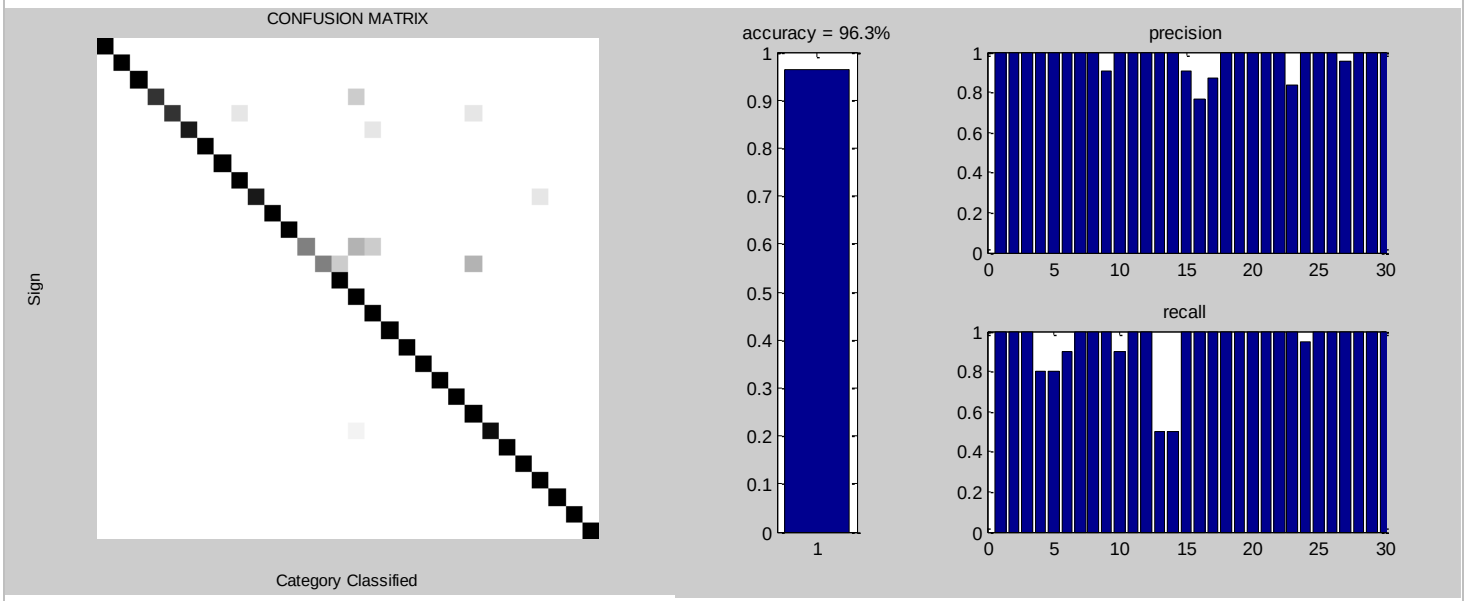
(Πλήρης κάλυψη του Ε3, μερική του Ε5 και Ε6)

Σ'αυτό το στάδιο καλούμαστε να κάνουμε κατάτμηση των νοημάτων σε τμήματα τα οποία παρουσιάζουν κοινά χαρακτηριστικά. Ουσιαστικά αυτό που θα ήταν το ιδανικό είναι να χωρίσουμε τα νοήματα σε υπομονάδες οι οποίες θα μπορούν να χαρακτηριστούν ως αντίστοιχα φωνήματα των φωνητικών σημάτων.

Αυτό που θα υλοποιήσουμε αρχικά είναι το εξής. Έχοντας στη διάθεσή μας 30 νοήματα θεωρούμε πιθανό αυτά τα νοήματα να δομούνται απο συνολικά 10 υπομονάδες(λιγες αλλα γενικές) . Έτσι για παράδειγμα το νόημα AIR μπορεί αν αποτελείται απο 3,4 περισσότερα ή λιγότερα φωνήματα. Για να το καταφέρουμε αυτό στη συνέχεια οργανώνουμε επαυξημένα δεδομένα που έχουμε εξάγει απο το δεύτερο μέρος το ένα κάτω απο το άλλο για όλα τα νοήματα δημιουργώντας έτσι έναν και μοναδικό πίνακα Nx10 οπου N είναι το συνολικό πλήθος όλων των πλαισίων όλων των νοημάτων. Στη συνέχεια εφαρμόζουμε έναν απλο kmeans αλγόριθμο αναζητώντας 10 κλάσεις. Τελικά κάθε πλαίσιο θα ανήκει σε μία κλάση και η κλάση αυτή χαρακτηρίζει την υπομονάδα στην οποία ανήκει το πλάισιο. Έτσι για κάθε νόημα μπορούμε να θεωρήσουμε οτι αυτό θα παράγεται με συνδυασμό αυτών των 10 υπομονάδων. Το σύστημα θα ξεχωρίσει απο μόνο του ποιές θα είναι οι υπομονάδες αυτές αλλα είναι προφανές απο τη λογική που ακολουθεί ο kmeans οτι για παράδειγμα αν η κίνηση του δεξιού χεριού είναι προς τα πάνω και το αριστερό χέρι είναι ακίνητο αυτό το πλαίσιο θα ανήκει σε διαφορετική υπομονάδα με το άν το δεξί χέρι κινείται προς τα κάτω μαζί με το αριστερό καθώς τα διανύσματα εισόδου θα είναι αρκετά μακριά και προφανώς θα ανήκουν σε διαφορετικές κλάσεις.

Φυσικά στη συνέχεια μπορούμε να συμπεριλάβουμε την κλάση στην οποία ανήκει το πλάισιο σαν ένα νέο χαρακτηριστικό εισόδου(επι κάποιο συντελεστή σηματνικότητας) και να δοκιμασουμε να εκπαιδεύσουμε το σύστημα με τα νέα επαυξημένα διανυσματα χαρακτηριστικών.

Το παραπάνω σκεπτικό υλοποιείται απο τα script add_subunits.m που δημιουργεί το διάνυσμα με την ταυτότητα των υπομονάδων και επαυξάνει το διάνυσμα των χαρακτηριστικών με αυτά αποθηκεύοντας τα νέα δεδομένα στο αρχείο data_with_subunits.mat. Μετά απ'αυτά μπορούμε να δούμε στο παρακάτω σχήμα τα αποτελέσματα γα το μοντέλο που έχουμε επιλέξει.

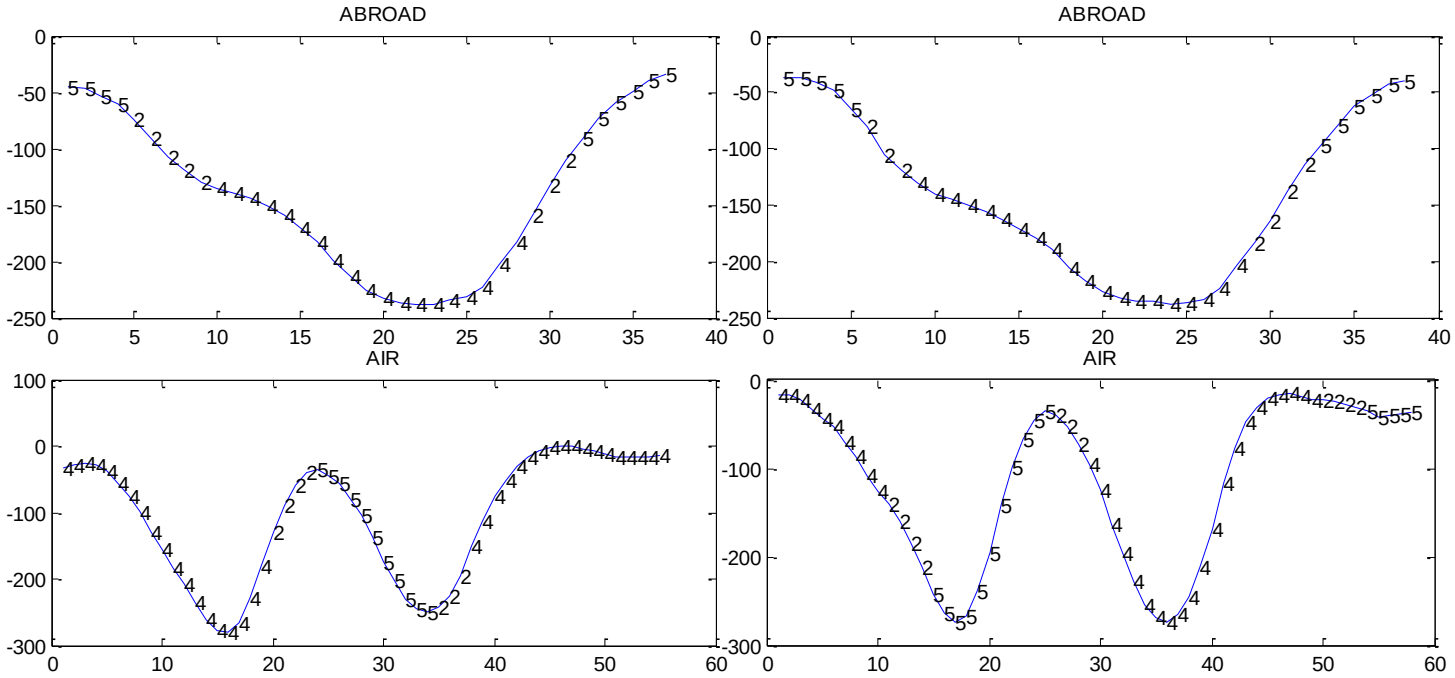


Σχήμα 13 : Αποτελέσματα μετά απο επάυξηση του διανύσματος χαρακτηριστικών με την ταυτότητα της κλάσης που ανήκει το κάθε πλαίσιο

Φυσικά πρέπει να αναφέρουμε οτι σε περίπτωση που έχουμε νέα άγνωστα δεδομένα θα πρέπει να μπορέσουμε να βρούμε σε ποιά κλάση υπομονάδας ανήκει το κάθε πλαίσιο. Για να γίνει αυτό θα πρέπει να έχουμε αποθηκευμένα τα κεντροειδή που προκύπτουν καθώς και την ταυτότητα του κάθε κεντροειδούς. Έπειτα με απλό αλγόριθμο k-NN μπορούμε να αποφανθούμε για την κλάση ως προς την υπομονάδα του κάθε πλαισίου.

Τα αποτελέσματα του παραπάνω πειράματος βλέπουμε οτι δεν είναι ικανοποιητικοτερα απ΄ότι εξήγαμε στο 2ο Μέρος. Ωστόσο θεωρούμε οτι αυξάνουμε τη γενίκευση του συστήματος και αυτό σημαίνει οτι οι μέσες τιμές των αποτελεσμάτων για νέα άγνωστα δεδομένα θα ήταν καλύτερες.

Πλέον μπορούμε επίσης να απεικονίσουμε τις υπομονάδες εμφανίζοντας για κάθε χαρακτηριστικό και την εξέλιξή του στον χρόνο την υπομονάδα στην οποία ανήκει. Συγκεκριμένα θα επιλέξουμε να εμφανίσουμε την μεταβολή της σχετικής x συντεταγμένης του δεξιού χεριού ενώ ταυτόχρονα θα εμφανίζουμε και τον αριθμό της υπομονάδας με χρήση της συνάρτησης text. Στο παρακάτω σχήμα παραθέτουμε τις πρώτες 2 επαναλήψεις 2 τυχαίων νοημάτων.



Σχήμα 14 : Απεικόνιση της συντεταγμένης x του δεξιού χεριού με ταυτόχρονη απεικόνιση της υπομονάδας στην οποία ανήκει η κίνηση

Όπως είναι αντιληπτό η χρήση μόλις 10 υπομονάδων για την παραγωγή 30 νοημάτων είναι μια πολύ κακή μοντελοποίηση. Ωστόσο οι υπομονάδες αυτές κρύβουν απλές πληροφορίες για τις κινήσεις και τις θέσεις των χεριών του νοηματιστή. Αντίστοιχα μπορούμε να χρησιμοποιήσουμε περισσότερες κλάσεις. Επίσης θα μπορούσαμε αντι για kmeans αλγόριθμο να χρησιμοποιούσαμε hmm και στη συνέχεια με χρήση του αλγορίθμου viterbi να βρίσκαμε την βέλτιστη ακολουθία καταστάσεων οπου η κάθε κατάσταση θα χαρακτήριζε μια υπομονάδα. Ωστόσο και ο απλός kmeans αλγόριθμος μπορεί να διαχωρίσει ικανοποιητικά το χώρο των χαρακτηριστικών σε διαφορετικές κλάσεις.

Οι υπομονάδες μπορούν να υπολογιστούν με χρήση του script subunits.m , να προστεθούν σαν νεο χαρακτηριστικό με το script add_subunits.m ενω η παραπάνω απεικόνιση μπορεί να πραγματοποιηθεί με το script plot_subunits.m.

Μέρος 5 : Ανασκόπιση ερωτημάτων

Είναι προφανές οτι στα πλαίσια της εργαστηριακής άσκησης αυτής δεν ακολουθήσαμε καθορισμένα βήματα καθώς σε κάθε μέρος έγιναν διάφορες δοκιμές εκπαίδευσης και ελέγχου ωστόσο μπορούμε πλέον να κάνουμε μια μικρή ανασκόπιση για κάθε υποερώτημα ξεχωριστά.

Ε1.
Σε αυτό το στάδιο επιλέξαμε να εξάγουμε τα ήδη υπάρχοντα χαρακτηριστικά αλλα και να προσθέσουμε κάποια επιπλέον. Όσον αφορά στα υπάρχοντα χαρακτηριστικά λάβαμε μόνο τις σχετικές διαφορές θέσεων των χεριών απο το κεφάλι ωστε τα αποτελέσματα να μην εξατώνται απο το που βρίσκεται ο νοηματιστής μέσα στο κάθε πλαίσιο αλλα απο τη σχετική κίνηση των χεριών του και του κεφαλιού. Επίσης κανονικοποιήσαμε τις εξαγόμενες τιμές με συντελεστές αντίστοιχους της ανάλυσης της εικόνας σε γραμμές και στήλες ωστε τα αποτελέσματα να μην εξαρτώνται απο τις διαστάσεις σε pixels του βίντεο και να λάβουμε ένα πιο γενικευμένο σύνολο χαρακτηριστικών το οποίο όπως είπαμε θα είναι κοινό και ανεξάρτητο απο τη θέση του νοηματισή αλλα και την ανάλυση του βίντεο. Αφου με αυτόν τον τρόπο μειώσαμε τη διάσταση χαρακτηριστικών κατα δύο χωρις ωστόσο να έχουμε σημαντική απώλεια πληροφορίας αλλα και να γενικεύσουμε τον τρόπο περιγραφής, στη συνέχεια προσθέσαμε νέα χαρακτηριστικά όπως τις διαφορές κατα x και κατα y των χεριών μεταξύ τους, και τις ταχύτητες κατα x και κατα y του κάθε χεριού. Έπειτα διοκιμάσαμε να προσθέσουμε εκ νέου χαρακτηριστικά αλλα η απόδοση του συστήματος άρχισε να μειώνεται. Τελικά θεωρούμε οτι με τον παραπάνω τρόπο καταφέραμε να περιγράψουμε με ικανοποιητικό τρόπο και να δημιουργήσουμε πλήρως περιγραφικά διανύσματα χαρακτηριστικών με τα οποία μπορούμε να εκπαιδεύσουμε το σύστημά μας, κάτι που εξάλλου επιβεβαιώνεται και απο τα αποτελέσματα.

Ε2.
Σε αυτό το στάδιο για τον αυτοματοποιημένο διαχωρισμό των πλαισίων δημιουργήσαμε μια συνάρτηση κριτηρίου με βάση τις οποίας έγινε ο διαχωρισμός σε ενεργές και μη ενεργές περιοχές. Συγκεκριμένα στη συνάρτηση αυτή που περιγράφει κάθε πλαίσιο αθροίζουμε τις manhatan αποστάσεις του κάθε χεριού απο τη θέση ισορροπίας καθώς και τις απόλυτες τιμές των ταχυτήτων των χεριών. Στη συνέχεια εμπειρικά επιλέγουμε ένα κατώφλι πάνω απο το οποίο θεωρούμε οτι τα πλαίσια βρίσκονται σε ενεργή περιοχή. Πλέον μπορούμε έυκολα να βρούμε την αρχή και το τέλος κάθε ενεργής περιοχής νοηματισμού. Τα αποτελέσματα έδειξαν οτι απο τις 150 συνολικά ενεργές περιοχές νοηματισμού βρέθηκαν με επιτυχία και πολύ καλό βαθμό ακρίβειας των ορίων των 145 περιοχών. Η εύρεση των περιοχών αυτών επιτυγχάνεται μέσω των του script segmentation.m και η εμφάνιση των διαφορών απο το script compare_is.m. Θεωρούμε οτι για τα δεδομένα αυτά επιτύχαμε μια αρκετά καλή κατάτμηση των βίντεο σε ενεργές περιοχές αν και αναμέναμε βάση του συστήματος να έχουμε πλήρη επιτυχία.

Ε3.
Σε αυτό το στάδιο έγινε προσπάθεια να χωρίσουμε τα νοήματα σε υπομονάδες. Με τον όρο υπομονάδες εννούμε κάτι αντίστοιχο των φωνημάτων στον προφορικό λόγο. Έτσι θα έπρεπε να βρούμε ομάδες οι οποίες παρουσιάζουν κοινά χαρακτηριστικά. Έτσι αποφασίσαμε να οργανώσουμε τα δεδομένα όλων των ομιλητών σε έναν και μοναδικό πίνακα μεγέθους Nx10 οπου 10 είναι το μέγεθος της διάστασης των διανυσματων χαρακτηριστικών περιγραφής και N είναι το συνολικό πλήθος των πλαισίων όλων των νοημάτων και όλων των επαναλήψεων. Στη συνέχεια εφαρμόσαμε έναν απλό kmeans αλγόριθμο αναζητώντας τόσες κλάσεις όσες και οι υπομονάδες με τις οποίες θεωρούμε οτι μπορούν να δημιουργηθούν τα 30 νοήματα. Αυτόβέβαια είναι κάτι αυθαίρετο καθώς με

μικρό αριθμό υπομονάδων μπορεί μια κλάση να χαρακτηρίζει το άν κινείται κάποιο χέρι ή οχι ενω με μεγαλύτερο πλήθος απο κλάσεις μπορούμε να μπούμε σε πιο εσωτερικές λεπτομέρειες. Στη συγκεκριμένη περίπτωση θεωρούμε οτι έχουμε 10 γενικευμένες υπομονάδες. Έτσι μετά το τέλος του kmeans το κάθε πλαίσιο θα ανήκει σε μια κλάση και φυσικά αναμένουμε οτι και γειτονικά πλαίσια θα ανήκουν στην ίδια ή σε γειτονικές κλάσεις εφόσον τα χαρακτηριστικά δεν έχουν απότομες μεταβολές απο πλαίσιο σε πλαίσιο. Τελικά μπορούμε να επαυξήσουμε το διάνυσμα χαρακτηριστικών ωστε κάθε πλαίσιο να περιγράφεται και απο την κλάση- υπομονάδα στην οποία ανήκει. Πραγματοποιώντας το παραπάνω είδαμε οτι η απόδοση του συστήματος δεν αυξήθηκε αλλα μειώθηκε κατα περίπου 2%. Αυτο δεν θεωρούμε οτι αποτελεί πρόβλημα καθώς μπορεί να χάνουμε απο ακρίβεια στο σύνολο δεδομένων των δύο νοηματιστών αλλα αυξάνουμε τη γενίκευση του συστήματος αναμένοντας να έχουμε μεγαλύτερη μέση απόδοση σε νέα άγνωστα δεδομένα. Επίσης απεικονίσαμε την εξέλιξη ενος χαρακτηριστικού στο χρόνο (δηλαδή απο πλαίσιο σε πλαισιο) και ταυτόχρονα την υπομονάδα απο την οποία χαρακτηρίζεται κάθε χρονική περιοχή κίονησης. Θεωρούμε οτι δεδομένου του γεγονότος οτι οι υπομονάδες της νοηματικής γλώσσας είναι άγνωστες σε αντίθεση με τα φωνήματα του προφορικού λόγου η παραπάνω διαδικασία είναι μια πρώτη αλλα πολύ γενική προσέγγιση της κατάτμησης του χώρου χαρακτηριστικών σε υποχώρους οπου ο καθένας περιγράφει μια υπομονάδα.

Ε5.

Όπως είδαμε μέσα απο την εργαστηριακή άσκηση αυτή έγιναν διάφορες δοκιμές και εφαρμόστηκαν διάφοροι αλγοριθμοι μη εκπαιδευόμενης μάθησης για την δημιουργία του τελικού συστήματος. Συγκεκριμένα για την εκπαίδευση του συστήματος απο τα χαρακτηριστικά εισόδου, για κάθε φώνημα εκπαιδεύσαμε ένα κρυφό μαρκοβιανό μοντέλο. Έγιναν δοκιμές για διαφορετικές τοπολογίες, διαφορετικό πλήθος καταστάσεων και γκαουσιανές που περιγράφουν την κάθε κατάσταση. Τελικά καταλήξαμε να επιλέξουμε τοπολογία που επιτρέπει ισοπίθανη μετάβαση απο κατάσταση σε κατάσταση έναντι του μοντέλου leftright που είδαμε οτι είναι αρκετά ισχυρό στην αναγνώριση φωνής. Ως αρχικές πριθανότητες θεωρήσαμε οτι το μοντέλο μπορεί να μεταβεί μόνο στην πρώτη κατάσταση καθώς αναγκάζουμε την πρώτη κατάσταση αυτή να είναι η θέση ηρεμίας-αρχική θέση του νοηματιστή. Βάση των πειραμάτων επιλέξαμε επίσης ως βέλτιστες τις 5 καταστάσεις και μια γκαουσιανή ανα κατάσταση οι οποίες περιγράφουν αρκετά καλά και γενικευμένα το πρόβλημα. Για την εκπαίδευση αρχικά χρησιμοποιήσαμε τα δεδομένα μόνο του πρώτου νοηματιστή αλλα τελικά εισαγάγαμε και τα δεδομένα του δεύτερου νοηματιστή, τουλάχιστον για τα νοήματα που δόθηκαν, λαμβάνοντας κατα κύριο λόγο το 80% των επαναλήψεων για κάθε κατηγορία ως δεδομένα εκπαίδευσης και το υπόλοιπο 20% ως τεστ (έγιναν δοκιμές και για 60-40).

Ως επιπλέον αλγόριθμοι μη εκαπαιδευόμενης μάθησης υπενθυμίζουμε οτι χρησιμοποιήθηκε ο απλός αλγόριθμος kmeans για τον διαχωρισμό των νοημάτων σε υπομονάδες με τρόπο που περιγράφεται στο Ε3 ενώ χρησιμοποιήθηκε και ένας επιπλέον αλγόριθμος για τον υπολογισμό των ενεργών και μη ενεργών περιοχών κατωφλιοποιώντας μια ευριστική συνάρτηση κριτηρίου (Ε2), ωστόσο για την εκπαίδευση του συστήματος αναγνώρισης των νοημάτων τα κρυφά μαρκοβιανά μοντέλα ήταν τα μοναδικά που μπορούν να περιγράψουν και να εκπαιδευτουν με βάση τη δυναμική του συστήματος, λαμβάνοντας υπόψιν την εξέλιξη του διανύσματος χαρακτηριστικών στον χρόνο.

Ε6.

Για την αξιολόγηση του συστήματος όπως περιγράψαμε αρκετά αναλυτικά στο 1ο Μέρος χρησιμοποιήσαμε έναν γενικευμένο πίνακα σύγχυσης (Confusion Matrix) οπου οπτικοποιήσαμε τα αποτελέσματα με κατάλληλο τρόπο επεξεργασίας του ενω επίσης μετρήθηκε η ακρίβεια ταξινόμησης αλλα και οι παράμετροι precision και recall. Στο πρώτο μέρος αναφέραμε αναλυτικά τον τρόπο που υπολογίζουμε αυτές τις παραμέτρους αλλα και τη σημαντικότητά τους. Επίσης σε κάθε περίπτωση

επιλέχθηκαν διάφοροι παράμετροι για την εκπαίδευση των μοντέλων και τελικά το πείραμα επαναλήφθηκε 10 φορές με τυχαίο διαμοιρασμό των δεδομένων σε training και testing set ωστε να εξάγουμε πιο αξιόπιστα αποτελέσματα. Αρχικά χρησιμοποιήσαμε τα δεδομένα μόνο του πρώτου νοηματιστή αλλα στη συνέχεια συμπεριλάβαμε και τον δεύτερο νοηματιστή τόσο στην εκπαίδευση όσο και στον έλεγχο. Μέσα απο την παρουσίαση και ανάλυση του κάθε μέρος ξεχωριστά έγινε αξιολόγηση σε κάθε βήμα και με διαφορετικές παραμέτρους και δεδομένα εισόδου.