ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ (MSc)
στα ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Κατάταξη Εκφωνημάτων Προφορικών Διαλόγων»

Γεώργιος Μπατιστάτος
Μ3110020

ΑΘΗΝΑ, ΦΕΒΡΟΥΑΡΙΟΣ 2013

ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ (MSc)
στα ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Κατάταξη Εκφωνημάτων Προφορικών Διαλόγων»

Γεώργιος Μπατιστάτος

Μ3110020

Επιβλέπων Καθηγητής: Ίων Ανδρουτσόπουλος
Εξωτερικός Κριτής: Θεόδωρος Καλαμπούκης

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΑΘΗΝΑ, ΦΕΒΡΟΥΑΡΙΟΣ 2013

ΠΡΟΛΟΓΟΣ

Η εκπόνηση της παρούσας διπλωματικής εργασίας πραγματοποιήθηκε στο πλαίσιο του Προγράμματος Μεταπτυχιακών Σπουδών στα Πληροφοριακά Συστήματα του Τμήματος Πληροφορικής του Οικονομικού Πανεπιστημίου Αθηνών. Διήρκησε από τον Ιούλιο του 2012 εώς τον Φεβρουάριο του 2013.

Επιθυμώ να εκφράσω τις ευχαριστίες μου σε όλους εκείνους που συνέβαλαν άμεσα ή έμμεσα στην ολοκλήρωση της διπλωματικής μου εργασίας και κατά συνέπεια των μεταπτυχιακών σπουδών μου.

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντά μου, Επίκουρο Καθηγητή Ίωνα Ανδρουτσόπουλο. Η υποστήριξη και διαθεσιμότητά του καθ΄ όλη τη διάρκεια της εκπόνησης της εργασίας αποτέλεσε σπουδαία βοήθεια, συμβάλλοντας ποικιλοτρόπως στην ολοκλήρωσή της, παρέχοντας μεταξύ άλλων πολύτιμες συμβουλές και καθοδήγηση όπου κρίθηκε αναγκαίο.

Θα ήθελα επίσης να ευχαριστήσω την εταιρία Omilia, χωρίς την βοήθεια της οποίας η παρούσα εργασία δεν θα μπορούσε να πραγματοποιηθεί, και ιδιαίτερα την κυρία Πέπη Σταυροπούλου. Η άμεση ανταπόκρισή της στις ανάγκες που προέκυπταν και η βαθιά γνώση του αντικειμένου και των δεδομένων ήταν καθοριστική.

Θα ήθελα ακόμα να ευχαριστήσω την ομάδα Επεξεργασίας Φυσικής Γλώσσας του Οικονομικού Πανεπιστημίου Αθηνών. Οι παρατηρήσεις και συμβουλές κάθε ξεχωριστού μέλους της οποίας βοήθησαν ουσιαστικά στη βελτίωση της εργασίας.

Τέλος, νιώθω την ανάγκη να ευχαριστήσω την οικογένειά μου για την αγάπη κατανόηση και υποστήριξή τους, χωρίς τις οποίες δεν θα ήταν δυνατή η ολοκλήρωση της παρούσας εργασίας.

ΠΕΡΙΛΗΨΗ

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η μελέτη, εφαρμογή και αξιολόγηση τεχνικών κατηγοριοποίησης κειμένου σε πραγματικά δεδομένα τα οποία προέκυψαν από την απομαγνητοφώνηση κλήσεων σε τηλεφωνικό κέντρο εταιρείας κινητής τηλεφωνίας που δραστηριοποιείται στην Ελλάδα. Τα δεδομένα ελήφθησαν με τη συμβολή εταιρείας που ασχολείται με συστήματα προφορικών διαλόγων η οποία συνεργάζεται με την εταιρεία κινητής τηλεφωνίας.

Έχει παρατηρηθεί ότι ένα συχνό ζήτημα που προκύπτει στα τηλεφωνικά κέντρα είναι η μειωμένη αποδοτικότητα και ποιότητα εξυπηρέτησης. Στο πλαίσιο της προσπάθειας βελτίωσης των υπηρεσιών τέτοιων κέντρων η παρούσα εργασία εξετάζει τη χρήση ευφυών τεχνικών για τη δρομολόγηση των κλήσεων. Πιο συγκεκριμένα, σκοπός είναι να εξετασθεί αν μέσω της χρήσης ενός ταξινομητή για την κατηγοριοποίηση των εκφωνημάτων μπορούν να επιτευχθούν καλύτερα ή εξίσου καλά αποτελέσματα όσον αφορά στην ορθή δρομολόγηση της κάθε κλήσης προς το καταλληλότερο τμήμα ή υπάλληλο για την εξυπηρέτησή της, συγκρίνοντας με την τρέχουσα προσέγγιση που χρησιμοποιείται στο τηλεφωνικό κέντρο. Η επίτευξη αυτού του στόχου επιφέρει σημαντικές συνέπειες στη λειτουργία ενός τηλεφωνικού κέντρου, όπως η μείωση του χρόνου αναμονής των πελατών και του κόστους λειτουργίας των κέντρων, η καλύτερη αξιοποίηση του χρόνου των υπαλλήλων και ενδεχομένως η αυτοματοποιημένη απάντηση απλών ερωτημάτων.

Στο υπάρχον σύστημα ο χρήστης-πελάτης είναι εκείνος που εκκινεί τη διαδικασία καλώντας το τηλεφωνικό κέντρο. Αρχικά, προτρέπεται να εκφράσει το αίτημά του, το οποίο μεταγράφεται με τη χρήση ενός συστήματος αναγνώρισης φωνής. Κατόπιν, το μεταγεγραμμένο εκφώνημα τροφοδοτείται σε μια χειρωνακτικά κατασκευασμένη γραμματική, η οποία αποφαίνεται για τη δρομολόγηση της κλήσης. Η απόφαση αυτή μεταφέρεται στο υποσύστημα δρομολόγησης που με τη σειρά του δρομολογεί την κλήση στον κατάλληλο υπάλληλο ή τμήμα. Είναι σαφές πως σε ένα τέτοιο σύστημα πραγματικού χρόνου η διαδικασία πρέπει να ολοκληρώνεται σε ένα εύλογο χρονικό διάστημα.

Τα αποτελέσματα που επιτυγχάνονται με τη χρήση γραμματικών είναι δυνατό να είναι ικανοποιητικά. Ωστόσο, η χρήση τους παρουσιάζει ορισμένα μειονεκτήματα. Η δημιουργία και ανανέωση μιας γραμματικής που επιτυγχάνει ικανοποιητικά αποτελέσματα χρειάζεται χρόνο, καθώς δεν μπορεί να γίνει εξολοκλήρου αυτόματα. Ένα μέρος της διαδικασίας πρέπει να γίνει με την επίβλεψη κάποιου ειδικού. Προφανώς, αυτό συνεπάγεται και αυξημένο κόστος, που ενδέχεται να είναι αρκετά υψηλότερο συγκριτικά με την χρήση ενός ταξινομητή που βασίζεται σε αλγορίθμους μηχανικής μάθησης. Δοθέντος ότι υπάρχουν αρκετά διαθέσιμα δεδομένα, ένας ταξινομητής αυτού του είδους μπορεί να εκπαιδευθεί και να ανανεωθεί αυτόματα και να επιτυγχάνει εξίσου καλά ή και καλύτερα αποτελέσματα.

Στο Κεφάλαιο 2 πραγματοποιείται μια επισκόπηση της έρευνας που αφορά τη δρομολόγηση κλήσεων με την παρουσίαση παρεμφερών συστημάτων δρομολόγησης και άλλων σχετικών στοιχείων. Η αρχική προσέγγιση δρομολόγησης κλήσεων στα τηλεφωνικά κέντρα εξυπηρέτησης ήταν εξ' ολοκλήρου χειροκίνητη, απαιτώντας από τον πελάτη να αλληλεπιδράσει με κάποιον υπάλληλο. Προσπάθειες που αναφέρονται στην βιβλιογραφία για την εισαγωγή νέων τεχνολογιών σε τέτοια συστήματα εξυπηρέτησης αφορούν την υποβοήθηση των υπαλλήλων στο να εξυπηρετούν αποδοτικότερα τις κλήσεις.

Οι πρώτες προσπάθειες που πραγματοποιήθηκαν για την αυτοματοποίηση της δρομολόγησης κλήσεων, και που ακόμα και σήμερα καλύπτουν σημαντικό μερίδιο της αγοράς, περιελάμβαναν την χρήση καταλόγων επιλογών που απαγγέλονται στους χρήστες. Αργότερα, εμφανίσθηκαν συστήματα που επιτρέπουν στους χρήστες να εκφράσουν ελεύθερα το αίτημά τους χρησιμοποιώντας προτροπές του τύπου «Πώς μπορώ να σας βοηθήσω;». Τα τελευταία είναι τα συστήματα που πραγματεύεται η παρούσα εργασία εστιάζοντας στο κομμάτι της κατηγοριοποίησης και όχι στην στρατηγική διαχείρισης διαλόγων, η οποία μπορεί να έχει εξίσου καθοριστικό ρόλο στην επιτυχία ενός τέτοιου συστήματος. Η κυριότερη προσέγγιση που απαντάται στη βιβλιογραφία για την κατηγοριοποίηση εκφωνημάτων σε τηλεφωνικά κέντρα είναι η διανυσματική αναπαράσταση των εκφωνημάτων και η χρήση τεχνικών ανάκτησης πληροφοριών ή / και αλγορίθμων μηχανικής μάθησης.

Προτού εξετασθούν τα διαθέσιμα δεδομένα και τα πειράματα που πραγματοποιήθηκαν στη διάρκεια της εργασίας, το Κεφάλαιο 3 αναφέρεται συνοπτικά στο απαραίτητο θεωρητικό υπόβαθρο εξετάζοντας τον ταξινομητή μεγίστης εντροπίας που χρησιμοποιήθηκε και καλύπτοντας κάποια στοιχεία που χρησιμοποιούνται στα πειράματα, όπως τα μέτρα αξιολόγησης, οι καμπύλες μάθησης και τα διαστήματα εμπιστοσύνης.

Στο Κεφάλαιο 4 περιγράφονται σε βάθος τα διαθέσιμα δεδομένα καθώς και οι διαφορετικοί τύποι κατηγοριών που χρησιμοποιήθηκαν για την επισημείωσή τους. Πρόκειται για μεταγεγραμμένα εκφωνήματα τα οποία προέρχονται από την αλληλεπίδραση χρηστών με το αυτοματοποιημένο σύστημα δρομολόγησης κλήσεων που λειτουργεί στο κέντρο εξυπηρέτησης πελατών της εταιρίας κινητής τηλεφωνίας.

Αρχικά, διαθέσιμο ήταν ένα σύνολο δεδομένων που αποτελούνταν από 75.268 καθαρά εκφωνήματα επισημειωμένα ώστε να ανήκουν το καθένα σε μια από 203 προκαθορισμένες κατηγορίες. Κατά την εκπόνηση όμως της εργασίας προέκυψαν διάφορες οπτικές του συγκεκριμένου σώματος εκφωνημάτων. Για να γίνει κατανοητό αυτό, σημειώνεται πως με τη συμβολή της εταιρίας που ασχολείται με συστήματα προφορικών διαλόγων τα μεν εκφωνήματα έγιναν διαθέσιμα τόσο στην καθαρή όσο και στη θορυβώδη μορφή τους, οι δε κατηγορίες που χρησιμοποιήθηκαν για την επισημείωση ήταν τελικά τριών διαφορετικών τύπων.

Διαθέσιμες από την αρχή ήταν οι προκαθορισμένες κατηγορίες. Το όνομα κάθε προκαθορισμένης κατηγορίας απαρτίζεται από δύο συνθετικά των 5 χαρακτήρων, που αναφέρονται ως πρόθημα και επίθημα. Κριτήρια για την επιλογή τους ήταν η κάλυψη των απαιτήσεων των πελατών και η αντιστοιχία με τις υπηρεσίες εξυπηρέτησης που προσφέρει η εταιρία κινητής τηλεφωνίας. Από το πρόθημα των προκαθορισμένων κατηγοριών προέκυψε ένα νέο είδος κατηγοριών οι οποίες ονομάσθηκαν ευρύτερες, ενώ η μελέτη των δεδομένων και των κατηγοριών κατά την εκπόνηση της εργασίας δημιούργησε την εντύπωση πως θα μπορούσε να ακολουθηθεί και μια διαφορετική προσέγγιση για την κατηγοριοποίηση, ασχέτως αν αυτή θα προοριζόταν για χρήση σε πραγματικό περιβάλλον. Το σκεπτικό ήταν πως η περαιτέρω «εξειδίκευση» των κατηγοριών θα μπορούσε να βελτιώσει τα

αποτελέσματα. Για να επιτευχθεί αυτό, αντί της χρήσης των προκαθορισμένων κατηγοριών, χρησιμοποιήθηκαν δύο διαφορετικοί τύποι κατηγοριών, οι κατηγορίες «services» και οι κατηγορίες «actions». Οι δύο αυτοί τύποι κατηγοριών δεν διαφέρουν δραστικά από τις προκαθορισμένες κατηγορίες. Οι κατηγορίες services περιλαμβάνουν κυρίως υπηρεσίες της εταιρίας κινητής τηλεφωνίας, ενώ οι κατηγορίες actions περιλαμβάνουν ενέργειες επί των υπηρεσιών. Σκοπός ήταν να προκύψει μια έκδοση του σώματος εκφωνημάτων που θα ήταν επισημειωμένη μόνο με τις κατηγορίες services και μια άλλη που θα ήταν επισημειωμένη μόνο με τις κατηγορίες actions. Τελικά, χρησιμοποιήθηκαν δύο διαφορετικοί ταξινομητές για την κατηγοριοποίηση, ο πρώτος για τα services και ο δεύτερος για τα actions και η σύνθεση των αποτελεσμάτων τους συγκρίθηκε με τα αντίστοιχα αποτελέσματα του μεμονωμένου ταξινομητή που εκπαιδεύθηκε στις αρχικές, προκαθορισμένες κατηγορίες.

Στο Κεφάλαιο 5 παρουσιάζονται αναλυτικά τα πειράματα που πραγματοποιήθηκαν. Τα αποτελέσματα που παρατηρήθηκαν ήταν ικανοποιητικά, ενώ σημαντικό είναι το γεγονός πως ήταν ελαφρώς καλύτερα από εκείνα της χειρωνακτικά δημιουργημένης γραμματικής που χρησιμοποιείται στο τηλεφωνικό κέντρο. Τα περισσότερα πειράματα πραγματοποιήθηκαν σε καθαρά εκφωνήματα, δηλαδή σε μεταγεγραμμένα εκφωνήματα που είχαν ελεγχθεί και διορθωθεί χειρωνακτικά. Σε πρώτη φάση πραγματοποιήθηκαν πειράματα με τις προεπιλεγμένες ρυθμίσεις του ταξινομητή, τόσο για τις προκαθορισμένες όσο και για τις ευρύτερες κατηγορίες. Ακολούθησε η δημιουργία μιας baseline μεθόδου, που κατατάσσει όλα τα εκφωνήματα στη συχνότερη κατηγορία. Τα αποτελέσματα της baseline μεθόδου ήταν πολύ χειρότερα, όπως θα ανέμενε κανείς. Τα επόμενα πειράματα αφορούσαν τη ρύθμιση μιας παραμέτρου του ταξινομητή που ελέγχει κατά πόσον ο ταξινομητής επιχειρεί να αποφύγει ή όχι την υπερεφαρμογή κατά το στάδιο της εκπαίδευσης, χωρίς όμως να παρατηρηθεί κάτι αξιοσημείωτο. Το ίδιο συνέβη και στην επόμενη σειρά πειραμάτων όπου χρησιμοποιήθηκαν *n*-γράμματα (ακολουθίες *n* λέξεων) ως ιδιότητες (features).

Στη συνέχεια, αξιοποιήθηκε η ιδιότητα του ταξινομητή να παρέχει για κάθε δεδομένο αξιολόγησης (εκφώνημα, στην περίπτωσή μας) μια κατανομή πιθανοτήτων, που δείχνει πόσο πιθανό θεωρεί ο ταξινομητής το δεδομένο να ανήκει σε κάθε μία

κατηγορία. Πραγματοποιήθηκε σειρά πειραμάτων, όπου επιτρεπόταν στον ταξινομητή να μην κατατάξει ένα εκφώνημα σε καμία κατηγορία (ώστε π.χ. να χειριστεί το εκφώνημα ένας άνθρωπος-τηλεφωνητής), όταν η βεβαιότητα του ταξινομητή ήταν κατώτερη ενός κατωφλίου. Εξετάσθηκε επίσης η χρήση δύο ταξινομητών έναντι ενός, αξιοποιώντας τις κατηγορίες services και actions και πάλι χωρίς σημαντικές διαφορές στα αποτελέσματα. Τέλος, συγκρίθηκε ο ταξινομητής με την γραμματική που χρησιμοποιείται στο τηλεφωνικό κέντρο.

Πειράματα πραγματοποιήθηκαν επίσης σε θορυβώδη εκφωνήματα, δηλαδή εκφωνήματα που προέκυψαν με αυτόματη μεταγραφή από ένα σύστημα αναγνώρισης φωνής, χωρίς χειρωνακτική διόρθωσή τους. Πραγματοποιήθηκαν διάφορες δοκιμές χρησιμοποιώντας μεταξύ άλλων διαφορετικού τύπου δεδομένα (μόνο καθαρά, μόνο θορυβώδη ή συνδυασμό) για την εκπαίδευση του ταξινομητή και την αξιολόγησή του.

Η εργασία ολοκληρώνεται με το Κεφάλαιο 6 όπου πραγματοποιείται ανακεφαλαίωση παρουσιάζοντας τα συμπεράσματα που προέκυψαν, τη συνεισφορά της εργασίας και ενδεχόμενες μελλοντικές κατευθύνσεις. Η βασική συνεισφορά της εργασίας έγκειται στην επίτευξη του σκοπού της, δηλαδή στην απόδειξη πως ένας αυτόματα εκπαιδευμένος ταξινομητής μπορεί να πετύχει εξίσου καλά αποτελέσματα με μια χειρωνακτικά κατασκευασμένη γραμματική.

Στις ενδεχόμενες μελλοντικές κατεθύνσεις για την περαιτέρω αξιοποίηση των δεδομένων σημειώνονται οι δοκιμές με διαφορετικούς αλγόριθμους μάθησης και διαφορετικές ιδιότητες, αλλά και η στενότερη σύνδεση του συστήματος δρομολόγησης των (μεταγεγραμμένων) εκφωνημάτων με το σύστημα αναγνώρισης φωνής. Για παράδειγμα, θα μπορούσε ενδεχομένως το σύστημα δρομολόγησης να χρησιμοποιεί και ιδιότητες που θα αντανακλούσαν τη βεβαιότητα αναγνώρισης (από το σύστημα αναγνώρισης φωνής) της κάθε λέξης.

SUMMARY

The subject of this MSc thesis is the study, application and evaluation of text classification techniques on a dataset containing transcribed customer utterances from the customer care call center of a telecommunications company operating in Greece. The dataset was obtained with the help of a company that provides spoken dialog systems, which has developed the system used by the telecommunications company.

It has been observed that a common issue in call centers is the reduced efficiency and low quality of service provided to the customers. In an effort to improve the service offered, this thesis examines the use of artificial intelligence techniques. More specifically, the purpose of this thesis is to examine whether or not using a classifier to categorize the transcribed customer utterances can lead to better (or at least as good) results, in terms of proper routing of each call to the appropriate department or employee, compared to the current approach used by the telecommunications company. A classifier of this kind could have a significant impact on the call center by reducing, for example, the waiting times of the customers, the operational cost of the call center, possibly also allowing simple requests to be answered automatically.

In the existing system the user-customer is the one who initiates the process by calling the call center. The system responds with a prompt like “How may I help you?”, and the user’s response is then transcribed using a speech recognition system. The transcribed utterance is then fed to a manually crafted grammar, which attempts to figure out the appropriate destination (e.g., department or employee) of the call. The decision of the grammar is then passed on to the routing subsystem, which in turn routes the call to the appropriate destination. Of course, with the system being real-time, the whole process must be completed in a reasonable time.

The results achieved using a manually developed grammar can actually be satisfactory. However, grammars have some clear disadvantages. Creating and updating a grammar can be time consuming, since the process typically requires the intervention of a natural language specialist. The cost of creating and maintaining a grammar may be much higher compared to the use of a classifier that is trained using

machine learning algorithms. Given that large datasets of user utterances are available from the call center, an automatically trained classifier may achieve equally good or even better results, compared to a manually crafted grammar.

In Chapter 2 an overview of relevant research on automated call routing is presented. The initial approach to call routing was entirely manual, requiring the customer to interact with a human operator. The first attempts to automate the process concerned the use of simple systems to assist the operators that interacted with the callers. Later attempts to automate call routing, which still account for a significant share of the market, used touch-tone menus. After calling, the customer has to wait and listen to the available options and then dial a number that corresponds to the desired action, in order to be routed to the corresponding agent and eventually be served. However, this approach is not completely satisfactory, as it often requires long waiting times and it involves repeating pre-recorded messages which may annoy the callers.

The next generation of automated call routing systems introduced “How may I help you?” prompts. In contrast to touch-tone menus, which are tedious to use, these more recent systems allow the customers to express their requests using natural language in response to an open prompt. These systems are the subject of this thesis. The thesis focuses on the classification of user utterances, rather than the overall dialogue management, even though the latter can be equally important.

The main approach found in the literature for utterance classification is the use of the vector space model. Each destination and input utterance can be represented as vectors. The destination is selected through the calculation of, for example, the cosine similarity between the utterance and the destination vectors. Another widely used approach is the use of machine learning techniques. Of course, there are some less conventional methods used too.

Before examining the available data and the experiments conducted, Chapter 3 presents the necessary theoretical background, providing information about the maximum entropy classifier used and covering some aspects of the experiments, such as the evaluation measures, the learning curves and the confidence intervals.

Chapter 4 puts forward a detailed description of the available data and the different types of classes used to annotate them. The dataset consists of transcribed utterances obtained through the interaction of the customers with the automated call routing system used in the call center. Initially, the available dataset consisted of 75,268 clean utterances annotated using 203 predefined classes. However, during the course of the thesis and with the help of the spoken dialogues company, the same utterances were made available in their noisy form, i.e., as they were output by the speech recognizers, without manual inspection and corrections. This permitted the examination of the data from different perspectives. Also, three different types of classes were used.

Initially, the only available classes were the predefined ones. Each predefined class consists of two components of 5 characters each, the prefix and suffix. The predefined classes were selected in accordance to customer requirements and the needs of the company operating the call center. A second kind of classes, named "wider categories", emerged by extracting the prefix of the predefined classes. Lastly, after getting accustomed with the data and the predefined classes, an attempt was made to produce a different inventory of classes, which might improve the results. The new classes reflected "actions" and "services". These two types of classes do not differ drastically from the predefined ones. While the services classes correspond to the services offered by the telecommunications company, the action classes correspond to possible actions related to the services.

The aim was to obtain a version of the dataset annotated only with the service classes and another version annotated only with the action classes. Then, two separate classifiers were used to classify each utterance as a service and action, respectively. The composition of their results was then compared to the corresponding result of the single classifier that was trained using the predefined classes.

Chapter 5 describes in detail the experiments that were carried out and the observed results. The results were satisfactory and, most importantly, they showed that an automatically trained classifier may achieve slightly better or at least as good results as those of the manually crafted grammar used in the call center.

Most experiments were performed on the clean dataset, i.e., utterances transcribed by a speech recognizer and then manually checked and corrected by humans. The first experiments were conducted using the default settings of the classifier for both the predefined and the wider classes. They were followed by the construction of a baseline, which classifies all the utterances in the most frequent class. As expected, the results of the baseline were not good. The next experiments involved trying different values of a parameter that controls the degree to which the classifier attempts to avoid over-fitting during training. The results showed no significant difference. There was also no significant difference in two next series of experiments that used *n*-grams (sequences of *n* words) as features and two distinct classifiers (one for services and one for actions) instead of a single one.

A different set of experiments were then conducted using the probabilities that the maximum entropy classifier assigns to the classes for each instance being classified. Instead of just returning a predicted class, the classifier can return a probability distribution showing how confident it is that the instance being classified belongs in each class. Experiments were conducted where the classifier was allowed not to classify an instance in any class, if its confidence was below a threshold (e.g., to allow the utterance to be handled by a human operator). Finally, an experimental comparison of the classifier and the grammar was performed, with the former showing slightly better results. A number of experiments were also performed using the noisy utterances. In particular, several tests were conducted including using different types of data (only clean, only noisy or a combination) to train and test the classifier.

The thesis concludes with Chapter 6 which contains a summary presenting the conclusions, the contribution and possible future directions of this thesis. The main contribution of the thesis is the achievement of its initial objective, namely to demonstrate that a classifier can achieve better or at least as good results as a manually crafted grammar.

The possibilities for future work include running tests using different machine learning algorithms and different features, as well as the closer association of the routing system with the speech recognition one. For example, the routing system could use features that would reflect the confidence of the speech recognition system for each transcribed word.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΡΟΛΟΓΟΣ 4
ΠΕΡΙΛΗΨΗ 5
SUMMARY 10
ΠΕΡΙΕΧΟΜΕΝΑ 15
1. ΕΙΣΑΓΩΓΗ 16
1.1. Αντικείμενο της Εργασίας 16
1.2. Περιγραφή του Προβλήματος 16
1.3. Περιγραφή του Υπάρχοντος Συστήματος 18
1.4. Στόχος 19
1.5. Δομή Εργασίας 19
2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ 21
3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ 26
3.1. Ταξινομητής 26
3.2. Καμπύλες μάθησης 26
3.3. Μετρικές 28
3.4. Διαστήματα εμπιστοσύνης 30
4. ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ 32
4.1. Εκφωνήματα 33
4.1.1. Καθαρά εκφωνήματα 33
4.1.2. Θορυβώδη εκφωνήματα 34
4.2. Κατηγορίες 35
4.2.1. Προκαθορισμένες κατηγορίες 35
4.2.2. Ευρύτερες κατηγορίες 38
4.2.3. Υποκατηγορίες 40
4.2.4. Κατηγορίες Services και Actions 41
5. ΠΕΙΡΑΜΑΤΑ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ 47
5.1. Καθαρά εκφωνήματα 47
5.1.1. Προεπιλεγμένες ρυθμίσεις ταξινομητή 47
5.1.2. Baseline (κατάταξη στην συχνότερη κατηγορία) 50
5.1.3. Παράμετρος Σίγμα 51
5.1.4. N-γράμματα 52
5.1.5. Κατώφλι κατάταξης 52
5.1.6. Σύγκριση του ταξινομητή με το προϋπάρχον σύστημα 53
5.1.7. Χρήση δύο ταξινομητών έναντι ενός 55
5.2. Θορυβώδη εκφωνήματα 55
6. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ 58
ΒΙΒΛΙΟΓΡΑΦΙΑ 59

1. ΕΙΣΑΓΩΓΗ

1.1. Αντικείμενο της Εργασίας

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η μελέτη, εφαρμογή και αξιολόγηση τεχνικών κατηγοριοποίησης κειμένου σε πραγματικά δεδομένα τα οποία προέκυψαν από την απομαγνητοφώνηση κλήσεων σε τηλεφωνικό κέντρο εταιρείας κινητής τηλεφωνίας που δραστηριοποιείται στην Ελλάδα. Τα δεδομένα ελήφθησαν με την συμβολή εταιρείας που ασχολείται με συστήματα προφορικών διαλόγων που συνεργάζεται με την εταιρεία κινητής τηλεφωνίας. Πρόκειται για περίπου 75.000 μεταγεγραμμένα εκφωνήματα τα οποία αξιοποιήθηκαν από διαφορετικές πλευρές για τη διεξαγωγή διαφόρων ειδών πειραμάτων. Οι ταξινομητές που χρησιμοποιούνται στα πειράματα της εργασίας είναι ταξινομητές μεγίστης εντροπίας. Σκοπός είναι να ερευνηθεί αν μέσω της κατηγοριοποίησης των εκφωνημάτων μπορούν να επιτευχθούν καλύτερα αποτελέσματα όσον αφορά στην ορθή δρομολόγηση της κάθε κλήσης προς το καταλληλότερο τμήμα ή υπάλληλο για την εξυπηρέτησή της, σε σύγκριση με τις μεθόδους που χρησιμοποιούνται ήδη στο τηλεφωνικό κέντρο.

1.2. Περιγραφή του Προβλήματος

Ένα τηλεφωνικό κέντρο δέχεται καθημερινά μεγάλο αριθμό κλήσεων, σημαντικό τμήμα των οποίων είναι πολύ απλά ερωτήματα ή αιτήματα. Τέτοιου τύπου αιτήματα, όπως ενδεικτικά η ενημέρωση για το υπόλοιπο του χρόνου ομιλίας ενός καρτοκινητού, θα μπορούσαν να εξυπηρετηθούν αυτόματα από υπάρχοντα συστήματα, χωρίς να χρειάζεται να απασχοληθεί υπάλληλος του τηλεφωνικού κέντρου. Επιπλέον, οι υπάλληλοι ενός τηλεφωνικού κέντρου δεν είναι δυνατό να έχουν κατάλληλη εξειδίκευση ώστε να απαντήσουν σε κάθε ερώτημα που ενδέχεται να προκύψει. Αφενός υπάρχουν εξειδικευμένοι υπάλληλοι, όπως εκείνοι του τεχνικού τμήματος και αφετέρου οι απαιτήσεις αλλάζουν ανά τακτά χρονικά διαστήματα με την προσθήκη νέων προϊόντων κλπ. Επομένως, συνήθως ένας πελάτης πρέπει να περάσει τουλάχιστον από δύο υπαλλήλους, έναν που θα δρομολογήσει το αίτημα, και

έναν που τελικά θα είναι ικανός να το αντιμετωπίσει. Αποτέλεσμα είναι να μην παρέχεται η καλύτερη δυνατή εμπειρία εξυπηρέτησης πρός τους πελάτες και το κόστος του τηλεφωνικού κέντρου να αυξάνεται λόγω απασχόλησης μεγάλου αριθμού υπαλλήλων.

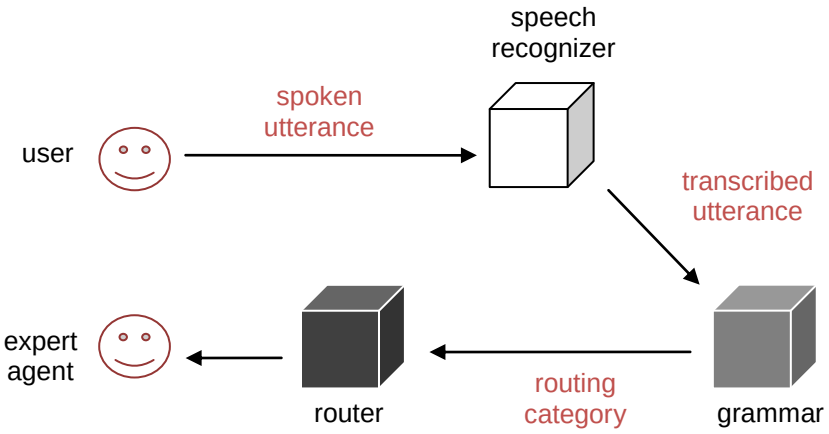
Από τα παραπάνω προκύπτει ότι είναι πολύ συχνό φαινόμενο να υπάρχει αρνητική εικόνα για τηλεφωνικά κέντρα με εμφανή την μη ικανοποίηση από την πλευρά των πελατών. Ταυτόχρονα, ένα θέμα που αντιμετωπίζουν οι ίδιες οι εταιρείες που παρέχουν τηλεφωνική εξυπηρέτηση είναι αυτό της μειωμένης αποδοτικότητας που επιτυγχάνεται.

Για λόγους σαν τους παραπάνω, οι οποίοι υποδηλώνουν τα περιθώρια βελτίωσης των υπηρεσιών που παρέχουν τα τηλεφωνικά κέντρα, γίνονται προσπάθειες αξιοποίησης νέων τεχνολογιών για την βελτίωσή τους. Ενδεικτικές τεχνικές που ήδη αξιοποιούνται είναι η αναγνώριση ομιλίας (speech recognition) ώστε να επιτευχθεί η αυτόματη δρομολόγηση των κλήσεων, η χρήση τεχνικών εξόρυξης γνώσης και επεξεργασίας φυσικής γλώσσας για την καλύτερη εξυπηρέτηση των πελατών, η καλύτερη εκπαίδευση των υπαλλήλων βάσει πρότυπων συμπεριφορών που έχει αναγνωρισθεί ότι επιφέρουν καλύτερα αποτελέσματα, είτε πρόκεται για εξυπηρέτηση σε εισερχόμενες κλήσεις είτε πρόκειται για πωλήσεις σε εξερχόμενες, και η αυτοματοποίηση διαδικασιών για την αύξηση της αποδοτικότητας των εκπροσώπων.

Μια τέτοιου είδους λύση είναι και η μέθοδος με την οποία η εταιρεία που ασχολείται με συστήματα προφορικών διαλόγων αυτοματοποιεί το πρόβλημα της δρομολόγησης των κλήσεων. Συγκεκριμένα, χρησιμοποιείται μια γραμματική για τον καθορισμό του τμήματος ή υπαλλήλου που θα εξυπηρετήσει την κάθε κλήση. Ωστόσο, παράλληλα με την χρήση γραμματικής, κρίθηκε σκόπιμη η μελέτη χρήσης ενός ταξινομητή, καθώς οι γραμματικές παρουσιάζουν κάποια μειονεκτήματα που αναφέρονται στην συνέχεια.

1.3. Περιγραφή του Υπάρχοντος Συστήματος

Λαμβάνοντας υπόψη πως το υπάρχον σύστημα είναι εμπορικό προϊόν πραγματικής εταιρείας, σκόπιμα παραλείπονται λεπτομέρειες για την εσωτερική του υλοποίηση που ούτως ή άλλως δεν αποτελούν το αντικείμενο της παρούσας διπλωματικής εργασίας. Έτσι λοιπόν το υπάρχον σύστημα παρουσιάζεται στην Εικόνα 1 σε αφαιρετική μορφή.



Εικόνα 1. Αφαιρετική αναπαράσταση της αρχιτεκτονικής του συστήματος.

Όπως φαίνεται, ο χρήστης είναι εκείνος που εκκινεί τη διαδικασία καλώντας το τηλεφωνικό κέντρο. Αρχικά, προτρέπεται να εκφράσει το αίτημά του, το οποίο μεταγράφεται με τη χρήση ενός συστήματος αναγνώρισης φωνής. Πλέον, το μεταγεγραμμένο εκφώνημα είναι δυνατό να τροφοδοτηθεί στη γραμματική, η οποία σύμφωνα με τους όρους του εκφωνήματος και τους κανόνες που την διέπουν αποφαίνεται για την δρομολόγηση της κλήσης. Η απόφαση αυτή μεταφέρεται στο router που με τη σειρά του δρομολογεί την κλήση στον κατάλληλο υπάλληλο ή τμήμα. Είναι σαφές πως σε ένα τέτοιο σύστημα πραγματικού χρόνου η διαδικασία πρέπει να ολοκληρώνεται σε ένα εύλογο χρονικό διάστημα.

Τα αποτελέσματα που επιτυγχάνονται με τη χρήση γραμματικών είναι δυνατό να είναι ικανοποιητικά. Ωστόσο, η χρήση τους παρουσιάζει ορισμένα μειονεκτήματα. Η δημιουργία και ανανέωση μιας γραμματικής που επιτυγχάνει ικανοποιητικά αποτελέσματα χρειάζεται χρόνο, καθώς δεν μπορεί να γίνει εξολοκλήρου αυτοματοποιημένα. Ένα μέρος της διαδικασίας πρέπει να γίνει με την επίβλεψη

κάποιου ειδικού. Προφανώς, αυτό συνεπάγεται και αυξημένο κόστος. Θα έλεγε λοιπόν κανείς πως το κόστος δημιουργίας και συντήρησης μιας γραμματικής είναι αρκετά υψηλότερο συγκριτικά με την χρήση ενός ταξινομητή, ο οποίος δοθέντος ότι υπάρχουν αρκετά διαθέσιμα δεδομένα μπορεί να δημιουργηθεί, εκπαιδευτεί και ανανεωθεί με αυτοματοποιημένο τρόπο και να επιτυγχάνει εξίσου καλά ή και καλύτερα αποτελέσματα.

1.4. Στόχος

Στόχος της εργασίας είναι η εξέταση και η παραμετροποίηση τεχνικών κατηγοριοποίησης που μπορούν εν δυνάμει να πετύχουν καλύτερα αποτελέσματα από την γραμματική που χρησιμοποιείται και πλεονεκτούν αυτής όπως αναφέρθηκε παραπάνω. Μέσω της κατηγοριοποίησης των εκφωνημάτων επιχειρείται να επιτευχθεί η ορθή δρομολόγηση της κάθε κλήσης προς το καταλληλότερο τμήμα ή υπάλληλο για την εξυπηρέτησή της. Η επίτευξη αυτού του στόχου επιφέρει σημαντικές συνέπειες στη λειτουργία ενός τηλεφωνικού κέντρου, συγκριτικά με τις συνήθεις πρακτικές. Η μείωση του χρόνου αναμονής των πελατών και του κόστους λειτουργίας των κέντρων, η καλύτερη αξιοποίηση του χρόνου των υπαλλήλων και ενδεχομένως η αυτοματοποιημένη απάντηση απλών ερωτημάτων συμπεριλαμβάνονται στα πλεονεκτήματα της προτεινόμενης μεθόδου.

1.5. Δομή Εργασίας

Ακολουθεί μια συνοπτική αναφορά στα κεφάλαια που έπονται.

Στο Κεφάλαιο 2 πραγματοποιείται μια επισκόπηση της έρευνας που αφορά την δρομολόγηση κλήσεων με την παρουσίαση παρεμφερών συστημάτων δρομολόγησης και άλλων σχετικών στοιχείων.

Το Κεφάλαιο 3 αναφέρεται συνοπτικά στο απαραίτητο θεωρητικό υπόβαθρο εξετάζοντας τον ταξινομητή και καλύπτοντας κάποια στοιχεία που χρησιμοποιούνται στα πειράματα.

Στο Κεφάλαιο 4 περιγράφονται σε βάθος τα διαθέσιμα δεδομένα καθώς και διαφορετικοί τύποι κατηγοριών που χρησιμοποιήθηκαν για την επισημείωσή τους.

Στο Κεφάλαιο 5 παρουσιάζονται αναλυτικά τα πειράματα που πραγματοποιήθηκαν.

Τέλος, στο Κεφάλαιο 6 πραγματοποιείται ανακεφαλαίωση παρουσιάζοντας τα συμπεράσματα που προέκυψαν, τη συνεισφορά της εργασίας και ενδεχόμενες μελλοντικές κατευθύνσεις.

2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ

Στο παρόν Κεφάλαιο παρουσιάζονται τα αποτελέσματα της βιβλιογραφικής μελέτης που πραγματοποιήθηκε στον τομέα της δρομολόγησης κλήσεων. Ένα ενδιαφέρον στοιχείο που προέκυψε κατά τη μελέτη ήταν η διαπίστωση πως η έρευνα στον συγκεκριμένο τομέα οδηγείται κυρίως από ερευνητές που απασχολούνται σε σημαντικές εταιρείες στον χώρο της Πληροφορικής και των Επικοινωνιών, όπως οι IBM, AT & T, Microsoft και άλλες.

Η αρχική προσέγγιση δρομολόγησης κλήσεων στα τηλεφωνικά κέντρα εξυπηρέτησης ήταν εξ' ολοκλήρου χειροκίνητη, απαιτώντας από τον πελάτη να αλληλεπιδράσει με κάποιον υπάλληλο. Προσπάθειες που αναφέρονται στην βιβλιογραφία για την εισαγωγή νέων τεχνολογιών σε τέτοια συστήματα εξυπηρέτησης αφορούν την υποβοήθηση των υπαλλήλων στο να εξυπηρετούν αποδοτικότερα τις κλήσεις αυτοματοποιώντας μερικώς τη διαδικασία. Στην εργασία των Lee και Chang [LEE] παρουσιάζεται ένα σύστημα υποβοήθησης των εκπροσώπων της Γενικής Διεύθυνση Τηλεπικοινωνιών της Ταϊβάν, το οποίο βασίζεται στο σκεπτικό ότι οι υπάλληλοι πετυχαίνουν καλή δρομολόγηση εφόσον δεν απαιτείται να γνωρίζουν κάθε δυνατή επιλογή, προτείνοντάς τους τις πιθανότερες. Για αυτό το λόγο, χρησιμοποιούνται λέξεις κλειδιά ανά κατηγορία και για κάθε κατηγορία υπολογίζεται ένας βαθμός ομοιότητας των λέξεων κλειδιών με το αίτημα του χρήστη. Οι Busemann κ.ά. [BUSE] πραγματοποίησαν μια μελέτη για την δρομολόγηση αιτημάτων σε μορφή e-mail γραμμένων στα γερμανικά, από τα οποία εξήχθησαν πληροφορίες όπως οι ρίζες των όρων, φράσεις που συνήθως αντιστοιχούσαν σε αρνήσεις, ερωτήσεις κλπ. Οι πληροφορίες αυτές αξιοποιήθηκαν σαν παράμετροι σε μια σειρά πειραμάτων όπου χρησιμοποιήθηκαν διαφορετικοί αλγόριθμοι ταξινόμησης ώστε να υποβοηθηθεί η απάντηση των συγκεκριμένων αιτημάτων.

Οι πρώτες προσπάθειες για την αυτοματοποίηση της δρομολόγησης κλήσεων, που ακόμα και σήμερα καλύπτουν σημαντικό μερίδιο της αγοράς, περιελάμβαναν τη χρήση καταλόγων επιλογών (touch tone menus) που απαγγέλονται στους χρήστες. Αφού ακούσουν τις διαθέσιμες επιλογές, οι χρήστες απαιτείται να επιλέξουν είτε φωνητικά είτε χρησιμοποιώντας το πληκτρολόγιό τους σαν συσκευή εισόδου. Η

συγκεκριμένη μέθοδος αφενός είναι χρονοβόρα και αφετέρου ορισμένες φορές απαιτείται να επαναληφθεί αρκετές φορές μέχρι ο πελάτης να φτάσει στο επίπεδο εκείνο του καταλόγου που περιέχει την απάντηση στο αίτημά του. Αποτέλεσμα των παραπάνω είναι το χαμηλό επίπεδο εξυπηρέτησης των χρηστών.

Σαν απάντηση στα προαναφερθέντα, εμφανίσθηκαν συστήματα ανοιχτών διαλόγων που επιτρέπουν στους χρήστες να εκφράσουν ελεύθερα το αίτημα τους χρησιμοποιώντας προτροπές του τύπου "How may I help you?". [GORIN] [RICC] [CARR] [DIFA]. Τα βασικά κοινά χαρακτηριστικά των προαναφερθέντων συστημάτων είναι ότι χρησιμοποιούν μια αρθρωτή αρχιτεκτονική και επεξεργάζονται τις κλήσεις σε μια σειρά βημάτων. Αρχικά, καταγράφεται και απομαγνητοφωνείται το εκφώνημα, και επιχειρείται η αναγνώρισή του. Στη συνέχεια, εφαρμόζονται μέθοδοι κατηγοριοποίησης ώστε να του αποδοθεί μια κατηγορία, και τελικά η κλήση δρομολογείται κατάλληλα. Τα τελευταία είναι τα συστήματα που πραγματεύεται η παρούσα εργασία, εστιάζοντας στο κομμάτι της κατηγοριοποίησης και όχι στην στρατηγική διαχείρισης διαλόγων, η οποία μπορεί να έχει εξίσου καθοριστικό ρόλο στην επιτυχία ενός τέτοιου συστήματος.

Η κυριότερη προσέγγιση που απαντάται στη βιβλιογραφία για την κατηγοριοποίηση εκφωνημάτων σε τηλεφωνικά κέντρα είναι η διανυσματική αναπαράσταση και η χρήση τεχνικών ανάκτησης πληροφοριών. Σύμφωνα με τη συγκεκριμένη προσέγγιση η δρομολόγηση κλήσεων είναι δυνατό να αναπαρασταθεί στον διανυσματικό χώρο. Χρησιμοποιώντας διάφορα στατιστικά στοιχεία, όπως οι εμφανίσεις των όρων, και συνήθως αφού τα δεδομένα περάσουν κάποιο στάδιο προεπεξεργασίας (π.χ. αφαίρεση stop-words), είναι δυνατή η κατασκευή διανυσμάτων για κάθε δυνατό προορισμό. Τα αιτήματα των πελατών, μετατρέπονται κι αυτά με τη σειρά τους σε διανύσματα και τελικά χρησιμοποιώντας κάποιο μέτρο όπως η ομοιότητα συνημιτόνου (cosine similarity) το κάθε αίτημα δρομολογείται στην κατηγορία - προορισμό με το μεγαλύτερο βαθμό ομοιότητας.

Την συγκεκριμένη προσέγγιση ακολουθούν οι Chu-Carrol και Carpenter [CHUC] για την δρομολόγηση σε 23 κατηγορίες, οι οποίοι μάλιστα υποστηρίζουν πως η χρήση μιας σιγμοειδούς συνάρτησης για την αντιστοίχηση της ομοιότητας συνημιτόνου σε

βαθμό εμπιστοσύνης βελτιώνει τα αποτελέσματα. Το σύστημά τους πετυχαίνει 93.8% ορθότητα δοθέντος όμως ότι δεν δρομολογεί το 10.2% των κλήσεων. Σημαντικό είναι επίσης το γεγονός πως το σύστημά τους αυτοματοποιεί την διαδικασία αποσαφήνισης σε περίπτωση που δεν μπορεί να αποφανθεί για κάποια κατηγορία.

Στο ίδιο μοτίβο οι Cox και Shahshahani [COX] εξετάζουν την χρήση δύο διαφορετικών μεθόδων βασισμένων σε διανύσματα. Η πρώτη αποσκοπεί στο ταίριασμα ενός νέου αιτήματος με έναν προορισμό - κατηγορία, ενώ η δεύτερη στο ταίριασμα με κάποιο παλιότερο αίτημα. Ακολουθώντας την δεύτερη πετυχαίνουν οι εσφαλμένες κατηγοριοποιήσεις να αντιστοιχούν μόλις στο 5.1% χρησιμοποιώντας Latent Semantic Analysis (LSA) για να εξομαλύνουν τα διανύσματα και στη συνέχεια Linear Discriminant Analysis (LDA) για να μειώσουν τις διαστάσεις τους και να αυξήσουν την διαφοροποίηση μεταξύ τους.

Οι Boye και Wiren [BOYE], σε αντίθεση με τις περισσότερες άλλες έρευνες στον τομέα της δρομολόγησης κλήσεων, προτείνουν οι κατηγορίες να είναι οργανωμένες σε επίπεδα ώστε να διευκολύνεται η συντήρησή τους και εξετάζουν τις συνέπειες που έχει αυτό στην κατηγοριοποίηση, στον σχεδιασμό των διαλόγων και την αξιολόγηση των αποτελεσμάτων. Το σύστημα της TeliaSonera που περιγράφουν υποστηρίζει δύο ειδών κατηγορίες, την semantic που αφορά το αντικείμενο του εκφωνήματος και μάλιστα αποτελείται από 3 συνθετικά (οικογένεια προϊόντος, είδος αιτήματος, ειδικότερο αντικείμενο) και το action που αφορά το αν θα γίνει η δρομολόγηση ή απαιτείται αποσαφήνιση. Ο ταξινομητής που χρησιμοποιείται είναι της εταιρείας Nuance και πετυχαίνει 85% ορθότητα ανάμεσα σε 123 κατηγορίες και 80% ορθότητα ανάμεσα σε 1500 κατηγορίες. Σημειώνεται πως τα αποτελέσματα αυτά αφορούν τις semantic κατηγορίες και μάλιστα αντιμετωπίζοντας τα 3 συνθετικά ενιαία. Σε επόμενο Κεφάλαιο περιγράφεται μια παρεμφερής προσπάθεια συνδυασμού δύο ταξινομητών - ειδών κατηγοριών που πραγματοποιήθηκε στο πλαίσιο της εργασίας.

Οι Fernandez, Ginzburg και Lappin [FERN] χρησιμοποιούν γνωστούς αλγορίθμους μηχανικής μάθησης για την κατηγοριοποίηση εκφωνημάτων, που ενώ δεν είναι απαραιτήτως συντακτικά ορθά ως προτάσεις, είναι ολοκληρωμένα νοηματικά, σε 15 κατηγορίες. Οι αλγόριθμοι που χρησιμοποιούν είναι οι J4.8, SLIPPER, TiMBL και

MaxEnt (μεγίστης εντροπίας) με τον τελευταίο να πετυχαίνει ελαφρώς καλύτερα αποτελέσματα.

Οι Chelba, Mahajan και Acero [CHEL1] έχουν διεξαγάγει μια σειρά πειραμάτων με σκοπό την σύγκριση μεταξύ διαφόρων προσεγγίσεων κατηγοριοποίησης κειμένου. Πιο συγκεκριμένα, συγκρίνουν την απόδοση δύο ειδών ταξινομητών: από την μία, των γλωσσικών μοντέλων n-γραμμάτων, που χρησιμοποιούνται ταυτόχρονα κατά την αναγνώριση ομιλίας αλλά και για την κατηγοριοποίηση των εκφωνημάτων, και από την άλλη του Naive Bayes και των ταξινομητών μεγίστης εντροπίας. Τα πειράματά τους διεξάγονται στο σώμα κειμένων (benchmark corpus) ATIS. Το συμπέρασμα στο οποίο καταλήγουν είναι ότι με με την χρήση μεθόδων μεγιστοποίησης πιθανοφάνειας, για την επιλογή των κατάλληλων παραμέτρων και την βελτιστοποίηση των βαρών των ταξινομητών, μπορούν να επιτύχουν σημαντική βελτίωση των αποτελεσμάτων. Παρόμοια αποτελέσματα παρουσιάζουν επίσης σε μια ακόμη εργασία τους [CHEL2].

Στη χρήση discriminative μεθόδων με σκοπό την σημαντική βελτίωση της ποιότητας των αποτελεσμάτων καταφεύγουν και οι Kuo κ.ά. [KUO1]. Στη συγκεκριμένη εργασία, ωστόσο, επιχειρούν να εκπαιδεύσουν ένα μοντέλο διανυσματικού χώρου που αντιπροσωπεύεται από έναν latent semantic πίνακα, χρησιμοποιώντας ως κριτήριο την ελαχιστοποίηση του σφάλματος κατηγοριοποίησης. Η προσέγγισή τους παράγει ικανοποιητικά αποτελέσματα ακόμα και για μικρό μέγεθος του συνόλου εκπαίδευσης, ενώ εκτός από features που υποδεικνύουν τη σχέση ενός εκφωνήματος με κάποια κατηγορία, η συγκεκριμένη μέθοδος χρησιμοποιεί και anti-features, τα οποία υποδεικνύουν αρνητική συσχέτιση με συγκεκριμένες κατηγορίες.

Τα περισσότερα συστήματα κατηγοριοποίησης εκφωνημάτων χρησιμοποιούν υπάρχοντα επισημειωμένα δεδομένα για να εκπαιδευθούν. Όμως η διαδικασία επισημείωσης είναι χρονοβόρα και έχει σημαντικό κόστος. Αξίζει να σημειωθούν προσπάθειες που στοχεύουν στη δημιουργία εξίσου καλών συστημάτων δρομολόγησης χωρίς την χρήση δεδομένων που επισημειώθηκαν γι’ αυτό τον σκοπό [WANG] [DIFA] [KUO2].

Η αξιολόγηση συστημάτων που στοχεύουν στην αυτοματοποιημένη δρομολόγηση των κλήσεων σε τηλεφωνικά κέντρα εξυπηρέτησης περιλαμβάνει τόσο ποιοτικά κριτήρια, τα οποία εν μέρει είναι υποκειμενικά, όσο και κριτήρια αποδοτικότητας, τα οποία μετρώνται με αντικειμενικό τρόπο. Σε γενικές γραμμές αξιολογείται τόσο η ικανοποίηση των χρηστών - πελατών, όσο και οι παράμετροι αλληλεπίδρασης με το σύστημα [GLASS][TURU] [WALK].

3. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

3.1. Ταξινομητής

Ο ταξινομητής που χρησιμοποιήθηκε για τα πειράματα αποτελεί λογισμικό της ομάδας επεξεργασίας φυσικής γλώσσας του Πανεπιστημίου Stanford [STNLP]. Πρόκειται για μια υλοποίηση σε Java ενός ταξινομητή μεγίστης εντροπίας (maximum entropy classifier). Μπορεί να χρησιμοποιηθεί για διαφορετικού τύπου δεδομένα αλλά ενδείκνυται για δεδομένα κειμένου αφού παρέχει αρκετές δυνατότητες επεξεργασίας των δεδομένων και εξαγωγής ιδιοτήτων (features) από συμβολοσειρές. Οι ιδιότητες αυτές που εξάγει καθορίζονται χρησιμοποιώντας παραμέτρους. Ως προεπιλογή χρησιμοποιεί δίτιμες (Boolean) ιδιότητες που προέκυψαν μέσω της αναπαράστασης σακιδίου λέξεων (Bag of Words), ενώ συσχετίζει την κάθε ιδιότητα θετικά ή αρνητικά με όλες τις κατηγορίες χρησιμοποιώντας βάρη που υπολογίζει με την L-BFGS quasi-Newton μέθοδο.

Πολύ σημαντική είναι η ταχύτητα με την οποία εκπαιδεύεται και αποκρίνεται ο ταξινομητής. Ειδικά ο χρόνος απόκρισης έχει ιδιαίτερη σημασία στη συγκεκριμένη περίπτωση καθώς μιλάμε για ένα σύστημα πραγματικού χρόνου. Από τα πειράματα που πραγματοποιήθηκαν προέκυψε πως ο μέσος χρόνος εκπαίδευσής του για το σύνολο των δεδομένων εκπαίδευσης (περίπου 70.000 εκφωνήματα) ήταν περίπου 10 λεπτά, ενώ ο μέσος χρόνο απόκρισης για τα δεδομένα αξιολόγησης ήταν 0,012 δευτερόλεπτα ανά εκφώνημα, με αποτέλεσμα να μην μειονεκτεί ούτε σε αυτό τον τομέα έναντι της γραμματικής (η οποία αποκρίνεται στιγμιαία).

3.2. Καμπύλες μάθησης

Η απεικόνιση των αποτελεσμάτων των περισσότερων πειραμάτων που διεξήχθησαν στο πλαίσιο της εργασίας πραγματοποιήθηκε χρησιμοποιώντας καμπύλες μάθησης. Η επιλογή αυτή βασίσθηκε στο γεγονός πως οι συγκεκριμένες γραφικές παραστάσεις διευκολύνουν την διεξαγωγή διαγνωστικών ελέγχων για τη λειτουργία ενός συστήματος [NG1].

Πιο συγκεκριμένα, οι καμπύλες μάθησης απεικονίζουν δύο μετρικές, το σφάλμα εκπαίδευσης, που αφορά την απόδοση ενός συστήματος κατά την κατηγοριοποίηση των ίδιων δεδομένων με τα οποία εκπαιδεύθηκε, και το σφάλμα αξιολόγησης, που περιγράφει την απόδοση του συστήματος κατά την κατηγοριοποίηση διαφορετικών δεδομένων από αυτά που χρησιμοποιήθηκαν για την εκπαίδευσή του. Οι μετρικές αυτές αποτυπώνονται χρησιμοποιώντας πολλά διαφορετικά υποσύνολα των δεδομένων εκπαίδευσης, ώστε τελικά να προκύψουν οι καμπύλες.

Βάσει των καμπυλών που σχηματίζονται είναι δυνατό να προσδιορισθεί αν το σύστημα πάσχει από υπερεφαρμογή (over-fitting), δηλαδή αν έχει προσαρμοστεί στις ιδιαιτερότητες των δεδομένων αξιολόγησης, ή περιορισμένο χώρο αναζήτησης (high bias), δηλαδή αδυνατεί να μάθει το ζητούμενο, και ανάλογα να ακολουθηθεί η κατάλληλη μέθοδος αντιμετώπισης.

Συγκεκριμένα, αν

- το σφάλμα εκπαίδευσης αυξάνεται απότομα όσο προστίθενται παραδείγματα εκπαίδευσης,
- το σφάλμα αξιολόγησης μειώνεται απότομα όσο προστίθενται παραδείγματα εκπαίδευσης και
- η διαφορά ανάμεσα στα δύο σφάλματα είναι πολύ μεγάλη,

τότε πιθανότατα το σύστημα πάσχει από υπερεφαρμογή οπότε ενδείκνυται η χρήση περισσότερων δεδομένων εκπαίδευσης, λιγότερων και καλύτερων ιδιοτήτων και ίσως η χρήση απλούστερου μοντέλου υποθέσεων.

Διαφορετικά, αν

- το σφάλμα εκπαίδευσης αυξάνεται ελάχιστα όσο προστίθενται παραδείγματα εκπαίδευσης,
- το σφάλμα αξιολόγησης μειώνεται ελάχιστα όσο προσθέτουμε παραδείγματα εκπαίδευσης και
- υπάρχει πολύ μικρή διαφορά ανάμεσα στα δύο σφάλματα,

τότε το σύστημα πάσχει από περιορισμένο χώρο αναζήτησης. Σε αυτή την περίπτωση η χρήση περισσότερων ιδιοτήτων ή - και ενός πιο περίπλοκου μοντέλου υποθέσεων θα βελτίωνε τα αποτελέσματα [IA1].

3.3.Μετρικές

Ως προεπιλογή, το λογισμικό που χρησιμοποιήθηκε για την διεξαγωγή των πειραμάτων υπολογίζει τα μεγέθη micro-averaged και macro-averaged F1.

Τα μεγέθη αυτά αποτελούν αρκετά καλά μέτρα καθώς το F-measure, όντας ο αρμονικός μέσος των δύο, λαμβάνει υπόψη του τόσο την ακρίβεια (precision) όσο και την ανάκληση (recall) των αποτελεσμάτων. Ωστόσο, όπως ειπώθηκε πριν, για τις καμπύλες μάθησης απαιτείται το σφάλμα, το οποίο προκύπτει από τον παρακάτω τύπο:

$$\texttt{Error rate} = 1 - \texttt{Accuracy}$$

Το ζήτημα αυτό ξεπερνάται αφού αποδεικνύεται πως σε περιπτώσεις όπως η συγκεκριμένη, όπου κάθε εκφώνημα είναι επισημειωμένο με μια μόνο κατηγορία και η απόφαση του ταξινομητή είναι επίσης μια μόνο κατηγορία (one-off classification), το micro-averaged F1 ισούται με το Accuracy. [MANN]

Απόδειξη

Κατά τα γνωστά:

$\texttt{tp}_i$: Τα εκφωνήματα που ο ταξινομητής ορθά κατατάσσει στην i-στή κατηγορία

$\texttt{tn}_i$: Τα εκφωνήματα που ορθά δεν κατατάσσει στην i-στή κατηγορία

$\texttt{fp}_i$: Τα εκφωνήματα που λανθασμένα κατατάσσει στην i-στή κατηγορία

$\texttt{fn}_i$: Τα εκφωνήματα που λανθασμένα δεν κατατάσσει στην i-στή κατηγορία

και έστω ότι παρακάτω $\texttt{tp}$, $\texttt{tn}$, $\texttt{fp}$, $\texttt{fn}$ είναι τα αθροίσματα των $\texttt{tp}_i$, $\texttt{tn}_i$, $\texttt{fp}_i$, $\texttt{fn}_i$, αντίστοιχα, όλων των κατηγοριών.

Εφόσον πρόκειται για one-off classification, κάθε fp_i απόφαση αντιστοιχεί σε μια fn_j απόφαση για μια ακριβώς άλλη (τη σωστή) κατηγορία και το αντίστροφο. Δηλαδή, κάθε εκφώνημα που κατατάσσεται σε μια κατηγορία χωρίς να ανήκει σε αυτή (αυξάνοντας το fp_i της κατηγορίας) τελικά δεν κατατάσσεται στην κατηγορία που θα έπρεπε (αυξάνοντας έτσι τον αριθμό των fn_i της τελευταίας). Έτσι λοιπόν $fp = fn$.

Είναι γνωστό επίσης πως τα micro-averaged Precision (P), Recall (R) και F1 ορίζονται ως εξής:

$$P = tp / (tp + fp)$$
$$R = tp / (tp + fn)$$
$$F1 = 2 * P * R / (P + R)$$
$$Accuracy = (tp + tn) / (tp + tn + fp + fn) \quad (0)$$

Εφόσον όμως $fp = fn$ είναι εμφανές πως $P = R = F1$. Οπότε στη συγκεκριμένη περίπτωση το micro-averaged F1 είναι ίσο με το micro-averaged Precision και το micro-averaged Recall, δηλαδή:

$$F1 = P = R = tp / (tp + fp) = tp / (tp + fn)$$

Έστω τώρα ότι $tp + fp = tp + fn = N$, όπου N ο αριθμός των κατατάξεων και άρα των εκφωνημάτων. Τότε:

$$F1 = tp / N \quad (1)$$

Σημειώνεται επίσης πως το άθροισμα $tp_i + tn_i + fp_i + fn_i$, για μία μόνο κατηγορία αυτή τη φορά, είναι ίσο με N, οπότε για c κατηγορίες:

$$tp + tn + fp + fn = c * N \quad (2)$$

Όταν εμφανίζεται ένα `tp` (όταν ένα εκφώνημα κατατάσσεται στη σωστή του κατηγορία), εμφανίζεται ένα tn_i για κάθε μία από τις υπόλοιπες `c-1` κατηγορίες, άρα συνολικά:

```
tn = (c - 1) * tp (3)
```

Αντικαθιστώντας τις σχέσεις `(2)(3)` στον τύπο `(0)` φαίνεται πως:

```
Accuracy = tp / N (4)
```

Τελικά, συνδυάζοντας τις σχέσεις `(1)(4)` αποδεικνύεται το αρχικό ζητούμενο:

```
Accuracy = F1
```

όταν τα Precision, Recall, F1 υπολογίζονται με micro-averaging και έχουμε one-off κατηγοριοποίηση, όπως παραπάνω.

3.4. Διαστήματα εμπιστοσύνης

Τα μεγέθη precision, recall και accuracy είναι αναλογίες (proportions) επιτυχιών σε ένα δείγμα μεγέθους n. Στην περίπτωση του precision, επιτυχίες είναι τα tp και n = tp + fp. Στην περίπτωση του recall, επιτυχίες είναι πάλι τα tp και n = tp + fn. Στην περίπτωση του accuracy, επιτυχίες είναι τα tp + tn και n = tp + tn + fp + fn = N. Επομένως, για να υπολογισθεί διάστημα εμπιστοσύνης για τέτοια μεγέθη, χρησιμοποιούμε τον τύπο διαστήματος εμπιστοσύνης για αναλογίες. Εφόσον έχουμε δείγματα μεγάλου μεγέθους (το n είναι πάντα μεγάλο), μπορεί να χρησιμοποιηθεί ο ακόλουθος τύπος (χρησιμοποιούμε α = 0.05):

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$$

Όπου $\hat{p}$ το μέγεθος (αναλογία) για το οποίο υπολογίζουμε διάστημα εμπιστοσύνης (δηλαδή precision, recall, accuracy), n το μέγεθος του δείγματος (όπως παραπάνω, ανά περίπτωση) και z ο συντελεστής της κανονικής κατανομής όπως προκύπτει από τους αντίστοιχους πίνακες.

4. ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ

Τα δεδομένα που χρησιμοποιήθηκαν για την διεξαγωγή των πειραμάτων της διπλωματικής εργασίας αποτελούν ιδιοκτησία εταιρείας κινητής τηλεφωνίας που δραστηριοποιείται στην Ελλάδα και χρησιμοποιήθηκαν εμπιστευτικά με τη συμβολή ελληνικής εταιρείας που ασχολείται με συστήματα προφορικών διαλόγων που συνεργάζεται με την πρώτη. Πρόκειται για μεταγεγραμμένα εκφωνήματα τα οποία προέρχονται από την αλληλεπίδραση χρηστών με το αυτοματοποιημένο σύστημα δρομολόγησης κλήσεων που λειτουργεί στο κέντρο εξυπηρέτησης πελατών της εταιρείας κινητής τηλεφωνίας.

Αρχικά, διαθέσιμο ήταν ένα σύνολο δεδομένων που αποτελούνταν από 75.268 καθαρά (χωρίς θόρυβο) μεταγεγραμμένα εκφωνήματα επισημειωμένα ώστε να ανήκουν σε μια από 203 κατηγορίες προκαθορισμένες από την εταιρεία. Κατά την εκπόνηση όμως της εργασίας προέκυψαν διάφορες οπτικές του συγκεκριμένου σώματος εκφωνημάτων. Τα μεν εκφωνήματα με τη συμβολή της εταιρείας έγιναν διαθέσιμα τόσο στην καθαρισμένη όσο και στη θορυβώδη μορφή τους, οι δε κατηγορίες που χρησιμοποιήθηκαν ήταν τελικά τριών διαφορετικών τύπων (βλ. Πίνακα 1).

	Εκφωνήματα	**Κατηγορίες**
1	Καθαρά	Προκαθορισμένες
2	Καθαρά	Ευρύτερες
3	Καθαρά	Κατηγορίες Services / Actions
4	Θορυβώδη	Προκαθορισμένες

Πίνακας 1. Οι οπτικές γωνίες από τις οποίες εξετάσθηκαν τα δεδομένα

Οι οπτικές γωνίες του Πίνακα 1 επεξηγούνται στα αμέσως επόμενα υποκεφάλαια, όπου εξετάζονται τα εκφωνήματα και οι κατηγορίες.

4.1. Εκφωνήματα

4.1.1. Καθαρά εκφωνήματα

Όπως ήδη αναφέρθηκε το σώμα εκφωνημάτων που ήταν αρχικά διαθέσιμο αποτελείται από καθαρά εκφωνήματα. Αυτό σημαίνει πως μέχρι να φθάσουν στην τρέχουσα καθαρή μορφή τους τα εκφωνήματα έχουν περάσει από τα εξής στάδια: Αρχικά από την αλληλεπίδραση των χρηστών με το αυτοματοποιημένο σύστημα δρομολόγησης προκύπτουν τα εκφωνήματα ως ηχητικά τεκμήρια, στη συνέχεια χρησιμοποιείται ένα σύστημα αναγνώρισης φωνής (speech recognizer) που επιχειρεί να μεταγράψει τα εκφωνήματα (να τα μετατρέψει σε γραπτή μορφή) και τελικά οι μεταγεγραμμένες μορφές των εκφωνημάτων υπόκεινται σε έλεγχο και διορθώσεις από ανθρώπους χρησιμοποιώντας τα διαθέσιμα ηχητικά τεκμήρια.

Στον Πίνακα 2 φαίνονται τρία αντιπροσωπευτικά παραδείγματα καθαρών εκφωνημάτων συνοδευόμενα από την κατηγορία στην οποία ανήκουν πριν εξετασθούν εκτενέστερα τόσο τα ίδια τα εκφωνήματα όσο και οι προκαθορισμένες κατηγορίες του συνόλου δεδομένων.

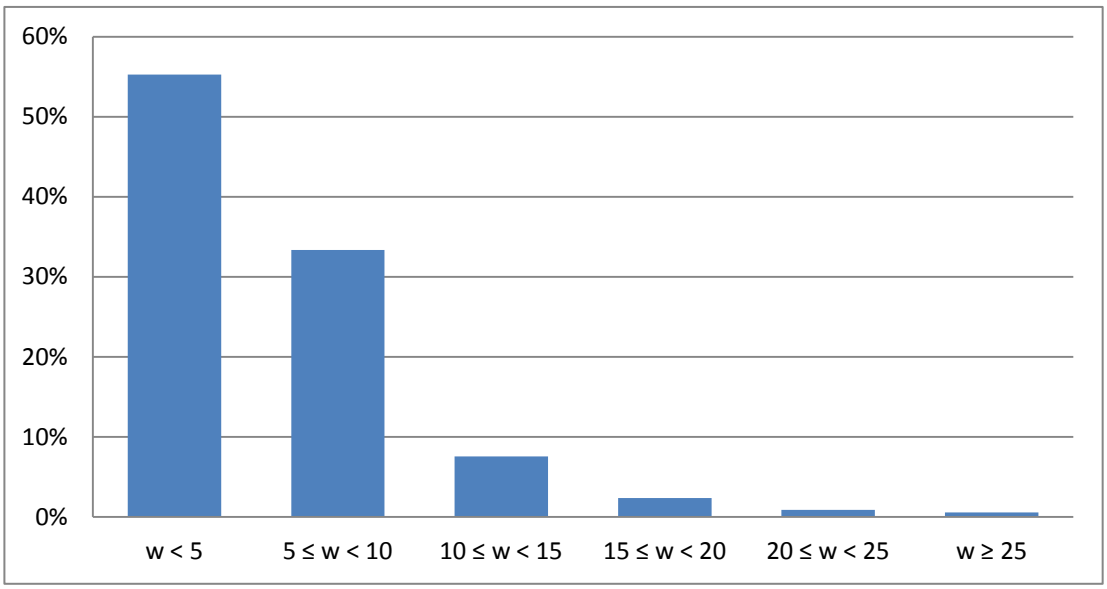
	Καθαρό Εκφώνημα (Διορθωμένο)	**Κατηγορία**
1	Θέμα λογαριασμού	BILLI_00000
2	Για τις ατιμολόγητες κλήσεις	BILLI_OTHER
3	Ήθελα να μάθω την επιδότησή μου	RETEN_00000

Πίνακας 2. Παραδείγματα καθαρών εκφωνημάτων

Τα εκφωνήματα περιέχουν λέξεις που μπορούν να βρεθούν σε λεξικό, καθώς και διάφορες άλλες «λέξεις» όπως το «εε» που προκύπτουν κατά την αποτύπωση των εκφράσεων που χρησιμοποιούνται συνήθως κατά τους προφορικούς διαλόγους.

Ένα άλλο στοιχείο που διακρίνεται είναι το γεγονός πως τα εκφωνήματα δεν έχουν κάποιο συγκεκριμένο μήκος (π.χ. αριθμό λέξεων). Το γεγονός αυτό φαίνεται και στην Εικόνα 2 όπου αποτυπώνεται η κατανομή του μήκους (σε λέξεις) των εκφωνημάτων.

Είναι φανερό πως τα περισσότερα εκφωνήματα είναι σύντομα. Ο μέσος όρος του μήκους τους είναι 5 λέξεις ενώ το 88.59% αυτών αποτελούνται από λιγότερες από 10 λέξεις. Αξίζει επίσης να σημειωθεί πως πολλά εκφωνήματα αποτελούνται από μια ή δύο λέξεις, με τα αντίστοιχα ποσοστά να είναι 17.16% και 18.46%. Το γεγονός αυτό είναι φυσιολογικό καθώς συνηθίζεται οι πελάτες που καλούν κέντρα εξυπηρέτησης πελατών να ενδιαφέρονται για κάποιο πολύ συγκεκριμένο θέμα - υπηρεσία που μπορεί να προσδιορισθεί με λίγες λέξεις.



Εικόνα 2. Κατανομή εκφωνημάτων βάσει μήκους σε αριθμό λέξεων

4.1.2. Θορυβώδη εκφωνήματα

Ίσως περισσότερο ενδιαφέρον από τα καθαρά εκφωνήματα παρουσιάζουν τα θορυβώδη (μη διορθωμένα), γιατί σε ένα σύστημα πραγματικού χρόνου θα ήταν αυτά που θα χρησιμοποιούνταν. Πρόκειται για τα αντίστοιχα εκφωνήματα που εμφανίζονται και στο αρχικό σύνολο δεδομένων, αυτή τη φορά όμως παραλείποντας το τελευταίο στάδιο, δηλαδή τον καθαρισμό και τη διόρθωση. Ουσιαστικά πρόκειται για τα εκφωνήματα όπως τα μεταγράφει το αυτόματο σύστημα αναγνώρισης ομιλίας. Στον Πίνακα 3 φαίνονται τα αντίστοιχα εκφωνήματα που παρουσιάσθηκαν στον Πίνακα 2, αυτή τη φορά στην θορυβώδη μορφή τους (όπως δηλαδή τα μετέγραψε το σύστημα αναγνώρισης φωνής).

	Θορυβώδες Εκφώνημα (όπως αποτυπώθηκε από το σύστημα αναγνώρισης φωνής)	**Κατηγορία**
1	Θέμα λογαριασμού	BILLI_00000
2	Για τις εε τιμολόγιο παίρνω κλήσεις	BILLI_OTHER
3	Ήθελα να μάθω τι επιδότησή μου	RETEN_00000

Πίνακας 3. Παραδείγματα θορυβωδών εκφωνημάτων

Ενώ όπως ειπώθηκε υπάρχει αντιστοιχία ανάμεσα στα δύο σύνολα δεδομένων, αυτό με τα θορυβώδη είναι λίγο μικρότερο σε μέγεθος. Το γεγονός αυτό οφείλεται εν μέρει στο ότι δεν ήταν διαθέσιμα όλα τα αντίστοιχα θορυβώδη εκφωνήματα και εν μέρει στις δυνατότητες του recognizer που σε ορισμένες περιπτώσεις όπου για παράδειγμα η ομιλία δεν ήταν καθαρή δεν αναγνωρίζει κάτι και άρα δεν εξάγει το αντίστοιχο θορυβώδες εκφώνημα. Πέρα λοιπόν από τα μη διαθέσιμα θορυβώδη εκφωνήματα που είναι περίπου 6.800, ακόμη 2.221 εκφωνήματα δεν αναγνωρίζει ο recognizer με αποτέλεσμα το θορυβώδες σύνολο δεδομένων να ανέρχεται στα 66.257 εκφωνήματα, δηλαδή περίπου 9.000 λιγότερα από το καθαρό σύνολο δεδομένων.

Ένα στατιστικό στοιχείο που είναι ιδιαίτερα ενδιαφέρον, είναι πως τα μισά περίπου (50,38%) εκφωνήματα που αναγνωρίζει ο recognizer αποτυπώνονται ορθά, ίδια δηλαδή με τα αντίστοιχα καθαρά, ενώ τα άλλα μισά (49,62%) διαφέρουν. Είναι, επομένως, αναμενόμενο τα αποτελέσματα στο θορυβώδες σύνολο δεδομένων να διαφέρουν από τα αποτελέσματα στο καθαρό σύνολο δεδομένων.

4.2. Κατηγορίες

4.2.1. Προκαθορισμένες κατηγορίες

Οι προκαθορισμένες κατηγορίες είναι 203 και επιλέχθηκαν από την εταιρεία που ασχολείται με συστήματα προφορικών διαλόγων σε συνεργασία με την εταιρεία κινητής τηλεφωνίας. Κριτήρια για την επιλογή τους ήταν η κάλυψη των απαιτήσεων των πελατών και η αντιστοιχία με τις υπηρεσίες εξυπηρέτησης που προσφέρει η εταιρεία κινητής τηλεφωνίας.

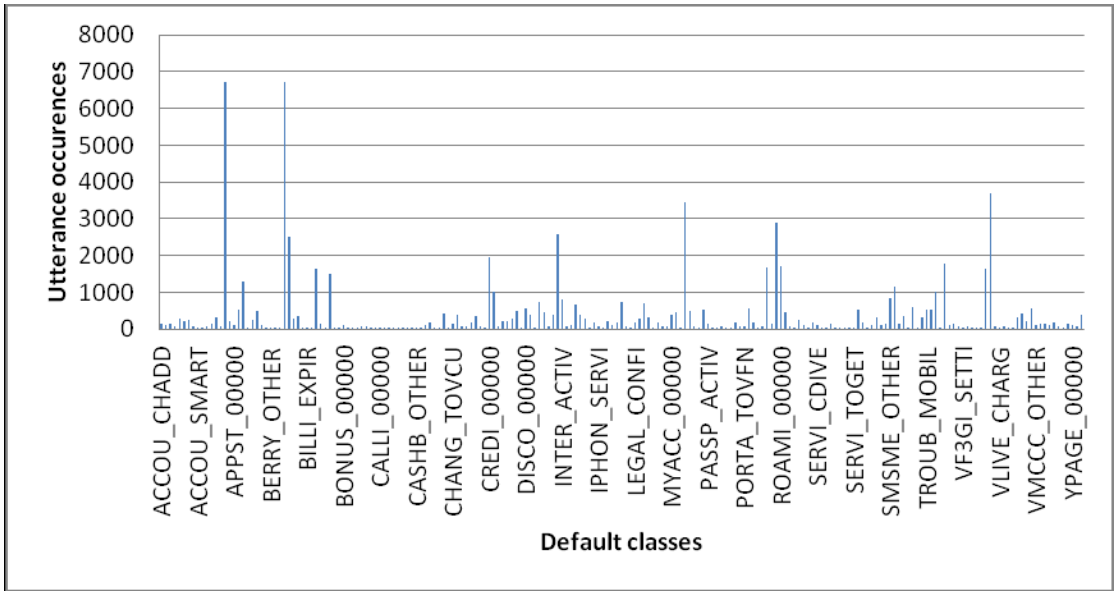
Το όνομα κάθε προκαθορισμένης κατηγορίας απαρτίζεται από δύο συνθετικά των 5 χαρακτήρων, που στο εξής θα αναφέρονται ως πρόθημα και επίθημα και συνδέονται με τον χαρακτήρα "_". Τα συνθετικά αυτά δεν είναι κάποιου συγκεκριμένου τύπου καθώς μεταξύ άλλων αναφέρονται σε υπηρεσίες, προϊόντα, ενέργειες κλπ. Ωστόσο, συνήθως το επίθημα έχει χαρακτήρα εξειδίκευσης, όπως φαίνεται από το τέταρτο παράδειγμα του Πίνακα 4.

ACTIV_00000	Ενεργοποίηση – γενικά
CHANG_CONTR	Αλλαγή συμβολαίου
INTER_00000	Ίντερνετ – γενικά
INTER_ACTIV	Ίντερνετ – ενεργοποίηση

Πίνακας 4. Χαρακτηριστικά παραδείγματα προκαθορισμένων κατηγοριών

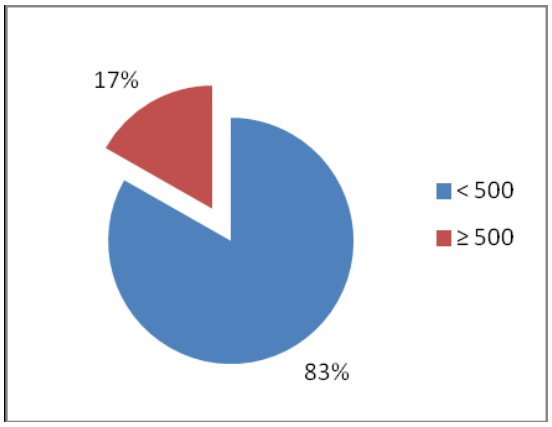
Ένα άλλο στοιχείο που προκύπτει από αυτά τα αντιπροσωπευτικά παραδείγματα είναι πως υπάρχουν συνθετικά που παρουσιάζονται τόσο ως προθήματα όσο και ως επιθήματα, όπως συμβαίνει με το «ACTIV» στο πρώτο και τέταρτο παράδειγμα του Πίνακα 4. Επιπλέον, φαίνεται πως το ίδιο πρόθημα αλλά και το ίδιο επίθημα μπορούν να εμφανίζονται σε περισσότερες από μια κατηγορίες, για παράδειγμα το τρίτο και τέταρτο εκφώνημα του Πίνακα 4 έχουν κοινό πρόθημα, ενώ το πρώτο και το τρίτο έχουν κοινό επίθημα.

Στην Εικόνα 3 παρακάτω φαίνεται η κατανομή των 75.268 εκφωνημάτων στις 203 προκαθορισμένες κατηγορίες. Τα δεδομένα είναι σε μεγάλο βαθμό άνισα κατανεμημένα με έναν περιορισμένο αριθμό κατηγοριών να συγκεντρώνουν την πλειοψηφία των εκφωνημάτων.

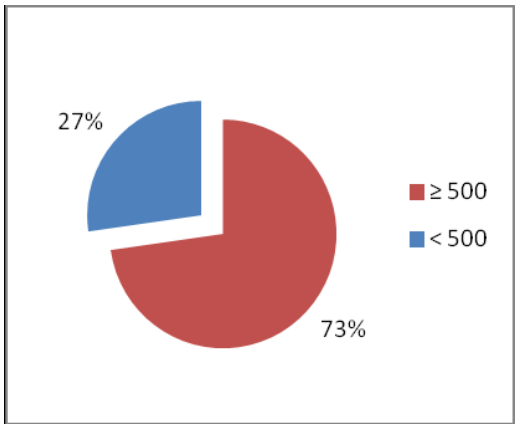


Εικόνα 3. Κατανομή των εκφωνημάτων ανά κατηγορία (προκαθορισμένες)

Συγκεκριμένα, 34 κατηγορίες συγκεντρώνουν περισσότερα από 500 εκφωνήματα, 18 περισσότερα από 1000 και μόλις 7 κατηγορίες περισσότερα από 2000, αντιστοιχώντας σε ποσοστό 16.5%, 8.74% και 3,4% επί του αριθμού των κατηγοριών (βλ. και Εικόνα 4).



Εικόνα 4. Ποσοστό κατηγοριών με ≥ 500 και < 500 εκφωνήματα

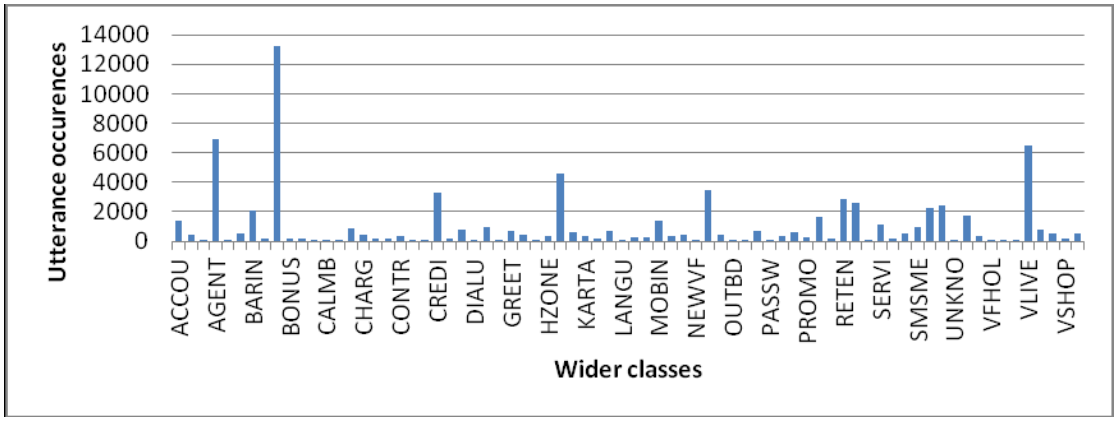


Εικόνα 5. Ποσοστό των εκφωνημάτων που ανήκουν σε κατηγορίες με ≥ 500 και < 500 εκφωνήματα

Τα στατιστικά αυτά στοιχεία (βλ. και Εικόνα 5) καταδεικνύουν μια παράμετρο που δυσχεραίνει την κατηγοριοποίηση των εκφωνημάτων, ιδιαίτερα εκείνων που ανήκουν στις μη συχνά εμφανιζόμενες κατηγορίες.

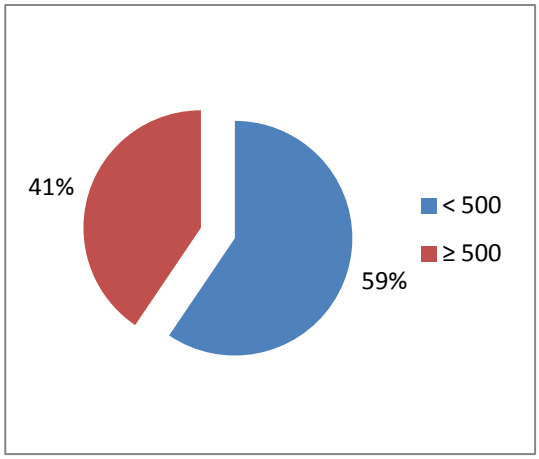
4.2.2. Ευρύτερες κατηγορίες

Οι ευρύτερες κατηγορίες είναι εκείνες που προκύπτουν εξετάζοντας μόνο το πρόθημα των προκαθορισμένων κατηγοριών.

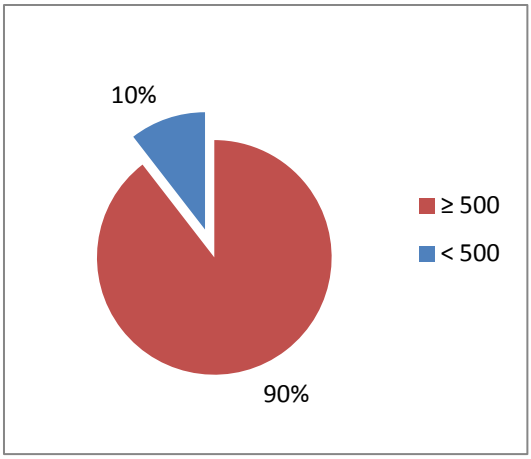


Εικόνα 6. Κατανομή των εκφωνημάτων ανά κατηγορία (ευρύτερες)

Στην Εικόνα 6, φαίνεται και πάλι το διάγραμμα που αναπαριστά την κατανομή των δεδομένων, αυτή τη φορά όμως μόνο στις 74 ευρύτερες κατηγορίες. Και σε αυτή την περίπτωση τα δεδομένα δεν είναι ομοιόμορφα κατανεμημένα, ωστόσο παρατηρείται μια σχετικά μεγαλύτερη ισορροπία συγκριτικά με τις προκαθορισμένες κατηγορίες. Πιο συγκεκριμένα, 44 ευρύτερες κατηγορίες, που αντιστοιχούν σε ποσοστό 59.46%, περιέχουν λιγότερα από 500 εκφωνήματα ενώ οι υπόλοιπες 30 (40.54%) περιέχουν 500 ή περισσότερα εκφωνήματα (βλ. Εικόνες 7 και 8). Από αυτές 14 περιέχουν ανάμεσα σε 500 και 1000 εκφωνήματα, 5 ανάμεσα σε 1000 και 2000 και 11 περισσότερα από 2000 εκφωνήματα.

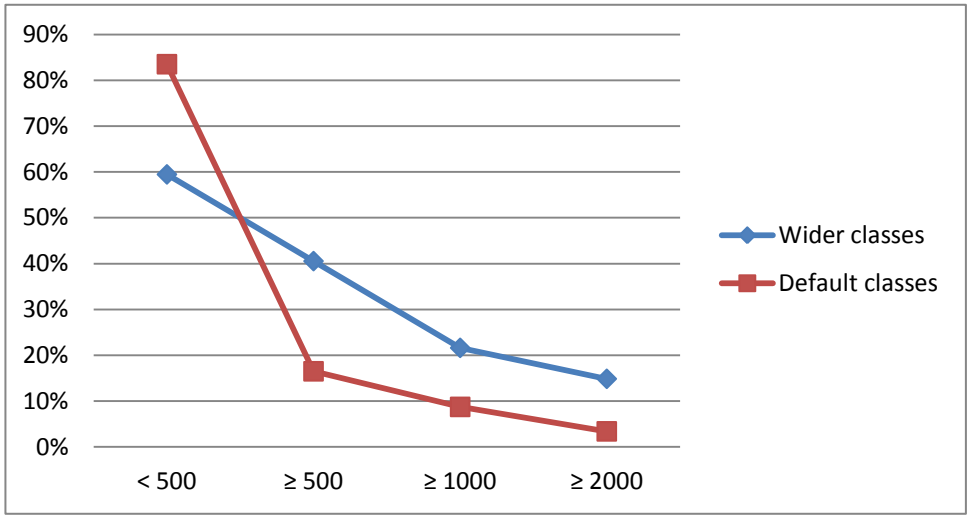


Εικόνα 7. Ποσοστό κατηγοριών με ≥ 500 και < 500 εκφωνήματα



Εικόνα 8. Ποσοστό των εκφωνημάτων που ανήκουν σε κατηγορίες με ≥ 500 και < 500 εκφωνήματα

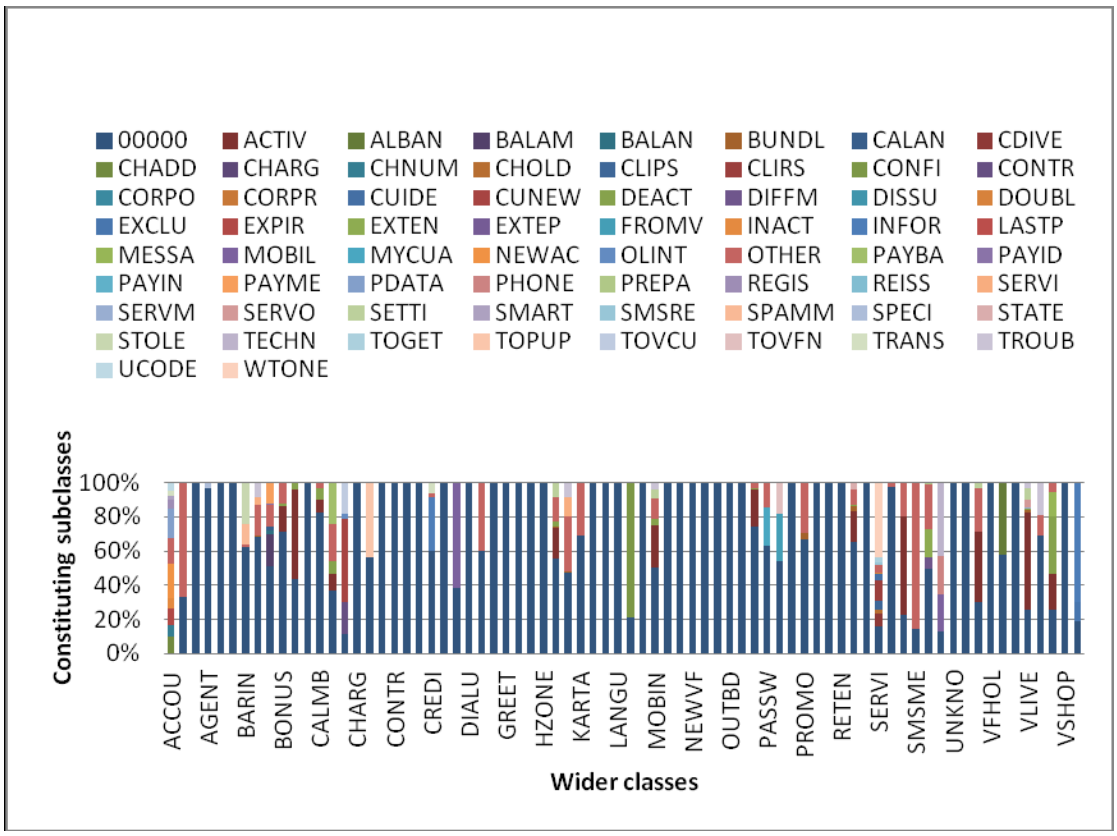
Στην Εικόνα 9 που ακολουθεί φανερώνεται τι ποσοστό των κατηγοριών έχουν αριθμό εκφωνημάτων που ανήκει σε μια από τις τέσσερις επιλεγμένες στάθμες τόσο για τις προκαθορισμένες όσο και για τις ευρύτερες κατηγορίες. Είναι εμφανές πως η καμπύλη των ευρύτερων κατηγοριών είναι ελαφρώς πιο ομαλή γεγονός που εξηγείται από τον αρκετά χαμηλότερο αριθμό τους.



Εικόνα 9. Συσχέτιση της κατανομής των εκφωνημάτων στις προκαθορισμένες και τις ευρύτερες κατηγορίες

4.2.3. Υποκατηγορίες

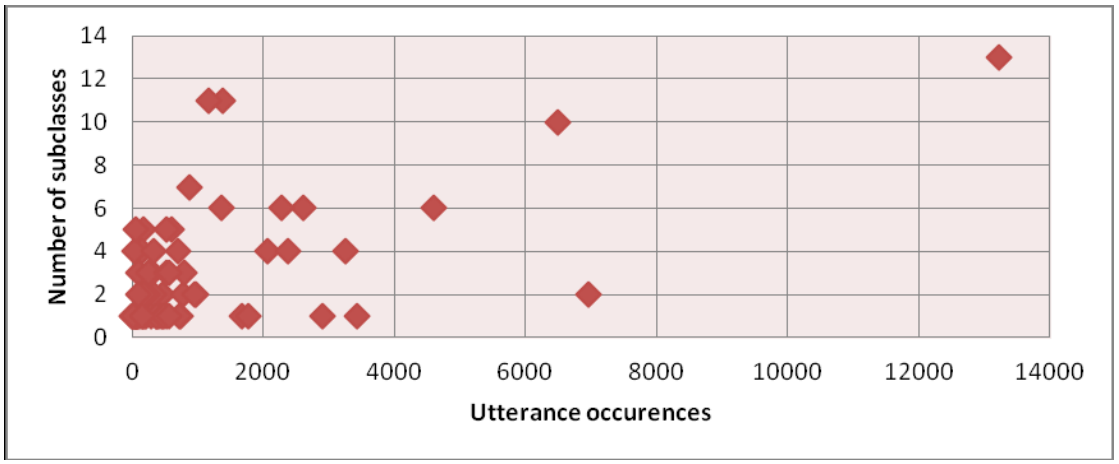
Εφόσον τα προθήματα των προκαθορισμένων κατηγοριών από μια οπτική γωνία είναι αντιληπτά σαν ευρύτερες κατηγορίες, τα αντίστοιχα επιθήματα μπορούν να θεωρηθούν σαν υποκατηγορίες αυτών και μάλιστα φτάνουν τις 66. Ο συγκεκριμένος τύπος κατηγοριών (οι υποκατηγορίες μόνες τους) δεν χρησιμοποιείται σε κάποιο από τα πειράματα. Ωστόσο, για λόγους πληρότητας εξήχθησαν κάποια ενδιαφέροντα στατιστικά στοιχεία και από αυτόν τον τύπο. Από τις 74 ευρύτερες κατηγορίες οι μισές ακριβώς έχουν μια μόνο υποκατηγορία (ουσιαστικά δεν έχουν), ενώ οι υπόλοιπες εμφανίζουν περισσότερες.



Εικόνα 10. Σύσταση των ευρύτερων κατηγοριών ως προς τις υποκατηγορίες που εφανίζουν

Στην Εικόνα 10 παραπάνω φαίνεται αναλυτικά η σύσταση της κάθε ευρύτερης κατηγορίας ως προς τις υποκατηγορίες που εμφανίζει, ενώ η Εικόνα 11 αποτυπώνει την συσχέτιση ανάμεσα στον αριθμό των εκφωνημάτων που ανήκουν σε μια

ευρύτερη κατηγορία (οριζόντιος άξονας) και τον αριθμό των υποκατηγοριών της συγκεκριμένης ευρύτερης κατηγορίας (κατακόρυφος άξονας).



Εικόνα 11. Συσχέτιση ανάμεσα στον αριθμό των εκφωνημάτων μιας ευρύτερης κατηγορίας και τον αριθμό των υποκατηγοριών της συγκεκριμένης ευρύτερης κατηγορίας

4.2.4. Κατηγορίες Services και Actions

Μέχρι στιγμής έχουν ήδη μελετηθεί διαφόρων τύπων κατηγορίες που είτε ήταν οι προκαθορισμένες είτε αποτελούσαν άμεσα παράγωγα αυτών. Ωστόσο, το γεγονός πως υπήρχαν προκαθορισμένες κατηγορίες για τα εκφωνήματα δεν σήμαινε αυτόματα πως η μόνη θεμιτή κατηγοριοποίηση στο πλαίσιο της εργασίας βασιζόταν σε αυτές. Η μελέτη των δεδομένων και των κατηγοριών κατά την εκπόνηση της εργασίας δημιούργησε την εντύπωση πως θα μπορούσε να ακολουθηθεί και μια διαφορετική προσέγγιση για την κατηγοριοποίηση, ασχέτως αν αυτή θα προοριζόταν για χρήση σε πραγματικό περιβάλλον. Το σκεπτικό ήταν πως η περαιτέρω «εξειδίκευση» των κατηγοριών θα μπορούσε να βελτιώσει τα αποτελέσματα.

Για να επιτευχθεί αυτό, αντί της χρήσης των προκαθορισμένων κατηγοριών, χρησιμοποιήθηκαν δύο διαφορετικοί τύποι κατηγοριών, οι κατηγορίες services και οι κατηγορίες actions. Οι δύο αυτοί τύποι κατηγοριών δεν διαφέρουν δραστικά από τις προκαθορισμένες κατηγορίες. Τα μεν services περιλαμβάνουν κυρίως υπηρεσίες της εταιρείας κινητής τηλεφωνίας, τα δε actions περιλαμβάνουν ενέργειες επί των υπηρεσιών. Αξίζει να σημειωθεί όμως πως ο διαχωρισμός δεν έγινε νοηματικά, καθώς

τα actions ουσιαστικά διαχωρίσθηκαν γιατί ήταν υπο-κατηγορίες οι οποίες μπορούσαν να βρεθούν σε πολλές από τις προκαθορισμένες κατηγορίες.

Για να γίνει κατανοητό το παραπάνω αναφέρεται ενδεικτικά πως το action ACTIV (ενεργοποίηση) μπορούσε να βρεθεί σε όλες τις παρακάτω προκαθορισμένες κατηγορίες:

1. ACTIV_00000	2. ACTIV_OTHER	3. BONUS_ACTIV	4. CALAL_ACTIV
5. CALMB_ACTIV	6. CASHB_ACTIV	7. INTER_ACTIV	8. MOBIN_ACTIV
9. PASSP_ACTIV	10. ROAMI_ACTIV	11. SIMCA_ACTIV	12. VF3GI_ACTIV
13. VLIVE_ACTIV	14. VOICE_ACTIV		

Πίνακας 5. Οι προκαθορισμένες κατηγορίες που περιείχαν το συνθετικό "ACTIV"

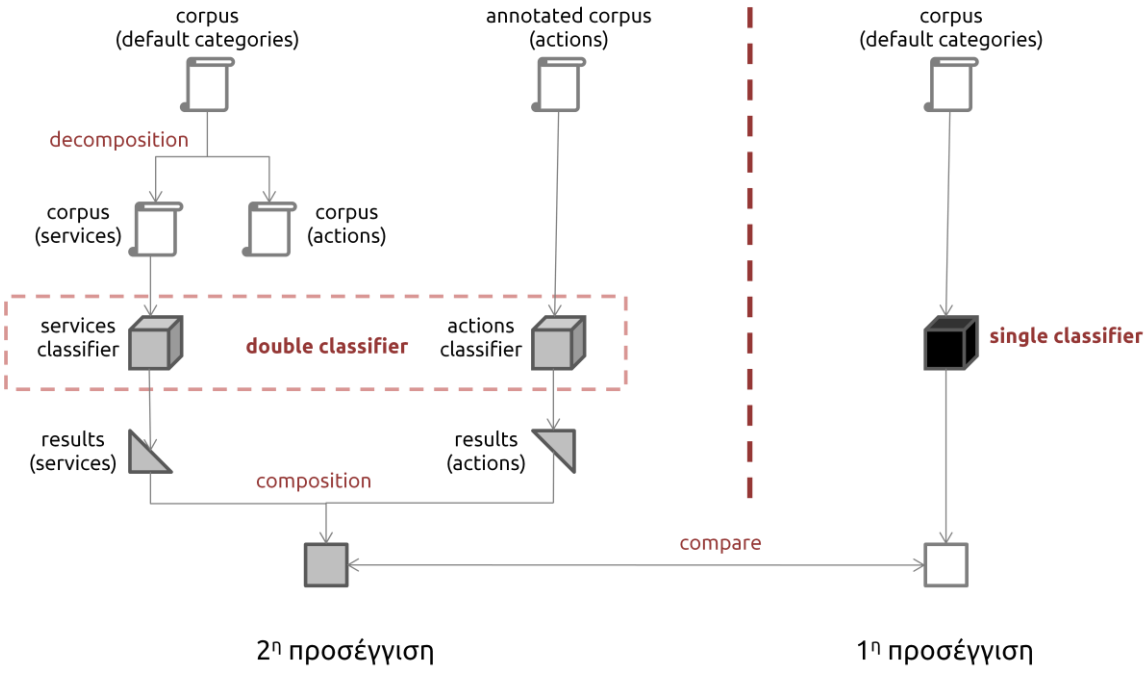
Έτσι τα actions που προέκυψαν είναι τα εξής:

1. ACTIVATE	2. AGENT	3. AGENTSPECIFIC	4. CHANGE
5. COMPLAINT	6. DEACTIVATE	7. GREETING	8. HELP
9. INFO	10. NOACTION	11. SETTINGS	12. TROUBLE
13. TROUBLETECH			

Πίνακας 6. Οι κατηγορίες actions που προέκυψαν

Οι κατηγορίες services προέκυψαν με διαφορετικό τρόπο. Πιο συγκεκριμένα, κατηγορίες services αποτελούν όλες εκείνες που προκύπτουν από την διαδικασία αποσύνθεσης που περιγράφεται αναλυτικά παρακάτω. Το πλήθος τους είναι τέτοιο που δεν επιτρέπει την παρουσίασή τους εδώ.

Σκοπός λοιπόν ήταν να προκύψει μια έκδοση του σώματος εκφωνημάτων που θα ήταν επισημειωμένη μόνο με τις κατηγορίες services και μια άλλη που θα ήταν επισημειωμένη μόνο με τις κατηγορίες actions. Τελικά, θα χρησιμοποιούνταν δύο διαφορετικοί ταξινομητές για την κατηγοριοποίηση, ο πρώτος για τα services και ο δεύτερος για τα actions και η σύνθεση των αποτελεσμάτων τους θα συγκρινόταν με τα αντίστοιχα αποτελέσματα του ταξινομητή που εκπαιδεύθηκε στις προκαθορισμένες κατηγορίες. Η παραπάνω διαδικασία φαίνεται στην Εικόνα 12 που ακολουθεί.



Εικόνα 12. Οι δύο διαφορετικές προσεγγίσεις κατηγοριοποίησης

Στην δεξιά στήλη φαίνεται η συνήθης διαδικασία με το σώμα εκφωνημάτων να περνά από έναν ταξινομητή, ενώ στην αριστερή στήλη φαίνεται η νέα προσέγγιση που εκμεταλλεύεται τις κατηγορίες actions και services. Στο μέσο της αριστερής στήλης αυτή τη φορά στη θέση του ενιαίου ταξινομητή βρίσκονται δύο ταξινομητές των οποίων τα αποτελέσματα συντίθενται και τελικά συγκρίνονται με τα αποτελέσματα του ταξινομητή της πρώτης προσέγγισης.

Κάθε ένας όμως από αυτούς τους δύο ταξινομητές χρειάζεται μια διαφορετικά επισημειωμένη έκδοση του σώματος εκφωνημάτων για να επιτελέσει το έργο του. Ο ταξινομητής των actions απαιτεί τα εκφωνήματα να είναι επισημειωμένα μόνο με τις κατηγορίες actions ενώ το ίδιο ισχύει για τον ταξινομητή των services. Οι δύο αυτές διαφορετικά επισημειωμένες εκδόσεις των εκφωνημάτων δημιουργήθηκαν με διαφορετικό τρόπο. Το σώμα εκφωνημάτων που επισημειώθηκε με τις κατηγορίες actions προέκυψε αρχικά χρησιμοποιώντας μια γραμματική για την επισημείωση και στη συνέχεια ελέγχοντας χειροκίνητα τα εκφωνήματα για τη διόρθωση των σφαλμάτων. Η διαδικασία αυτή ήταν ιδιαίτερα χρονοβόρα. Αντιθέτως, η έκδοση των εκφωνημάτων που επισημειώθηκε με τις κατηγορίες services πραγματοποιήθηκε με αυτοματοποιημένο τρόπο από μια διαδικασία αποσύνθεσης των επισημειωμένων προκαθορισμένων κατηγοριών αντίστοιχη με την εξαγωγή των ευρύτερων

κατηγοριών. Από τα εκφωνήματα αφαιρέθηκαν τα τμήματα εκείνα των κατηγοριών (είτε το πρόθημα είτε το επίθημα) που αντιστοιχούσαν σε actions και το υπόλοιπο αντιστοιχούσε σε μια κατηγορία service.

Κατά τον διαχωρισμό αυτό παρατηρήθηκαν περιπτώσεις που απαιτούσαν διαφορετική αντιμετώπιση, παραδείγματα των οποίων φαίνονται στον Πίνακα 7, και αντιμετωπίσθηκαν ως εξής:

DEFAULT CATEGORY	**SERVICE**	**ACTION**
ACCOU_CHADD	ACCOU_CHADD	NOACTION
AGENT_SPECI	00000	AGENTSPECIFIC
TROUB_TECHN	00000	TROUBLETECH
CHANG_CUNEW	CUNEW	CHANGE
CREDI_INFOR	CREDI	INFO
HELPR_00000	00000	HELP

Πίνακας 7. Οι ιδιαίτερες περιπτώσεις που παρατηρήθηκαν κατά την αποσύνθεση των προκαθορισμένων κατηγοριών

Από τις περιπτώσεις που φαίνονται ιδιαίτερη αναφορά αξίζει στη δεύτερη και τρίτη, που αφορούν δύο προκαθορισμένες κατηγορίες, τις AGENT_SPECI (συγκεκριμένος εκπρόσωπος) και TROUB_TECHN (τεχνικό πρόβλημα). Και στις δύο περιπτώσεις, ολόκληρη η προκαθορισμένη κατηγορία αντιστοιχεί σε ένα action με αποτέλεσμα σαν service να τίθεται το 00000 που υποδηλώνει πως δεν ανήκει κάπου συγκεκριμένα. Ενδιαφέρον παρουσιάζει επίσης το γεγονός πως ακολουθώντας τη συγκεκριμένη τακτική, το μήκος των κατηγοριών services δεν είναι σταθερό. Σε αντίθεση με τα υπόλοιπα παραδείγματα, η κατηγορία action του πρώτου είναι η NOACTION με αποτέλεσμα το αντίστοιχο service να αποτελείται από δύο συνθετικά. Η ανομοιομορφία αυτή θα μπορούσε να προκαλέσει προβλήματα στη συνέχεια.

Προτού πραγματοποιηθεί η τελική σύγκριση ανάμεσα σε έναν συμβατικό ταξινομητή και στον συνδυασμό των αποτελεσμάτων των δύο επιμέρους ταξινομητών (ένας για services κι ένας για actions) απαιτείται η αντίστροφη διαδικασία, δηλαδή η σύνθεση των αποτελεσμάτων των δύο ταξινομητών. Οι κυριότερες πιθανές έξοδοι των δύο επιμέρους ταξινομητών φαίνονται στον Πίνακα 8 και συνοδεύονται από το αποτέλεσμα που προκύπτει από την σύνθεσή τους. Σημειώνεται πως στον

συγκεκριμένο Πίνακα, οι όροι Service και Action χρησιμοποιούνται σαν μεταβλητές που αναφέρονται στις αντίστοιχες κατηγορίες. Οι παρενθέσεις που συνοδεύουν τον όρο Service χρησιμοποιούνται για να προσδιορίσουν το μήκος του, με τη λέξη single να υποδηλώνει πως αποτελείται από ένα συνθετικό και τη λέξη double από δύο, καθώς οι δύο περιπτώσεις απαιτούν διαφορετική αντιμετώπιση.

	1ος ΤΑΞΙΝΟΜΗΤΗΣ (SERVICES)	**2ος ΤΑΞΙΝΟΜΗΤΗΣ (ACTIONS)**	**ΑΠΟΤΕΛΕΣΜΑ**
1	Service	NOACTION	Service
2	Service (single)	Action	Service_Action / Action_Service
3	00000	AGENTSPECIFIC / TROUBLETECH	AGENT_SPECI / TROUB_TECHN
4	Service (single)	AGENTSPECIFIC / TROUBLETECH	?
5	Service (double)	Action	?

Πίνακας 8. Η μεθοδολογία της σύνθεσης δύο αποτελεσμάτων

Ο Πίνακας 9 που ακολουθεί περιέχει παραδείγματα που αντιστοιχούν στις περιπτώσεις του Πίνακα 8. Όπου προηγουμένως υπήρχε μια μεταβλητή πλέον έχει αντικατασταθεί από μια πραγματική κατηγορία:

	SERVICE	**ACTION**	**ΣΥΝΘΕΣΗ**
1	VLIVE_00000	NOACTION	VLIVE_00000
2	CREDI	INFO	CREDI_INFOR
3	00000	AGENTSPECIFIC	AGENT_SPECI
4	PHONE	TROUBLETECH	?
5	BILLI_BALAM	INFO	?

Πίνακας 9. Παραδείγματα σύνθεσης

Η περίπτωση 1 είναι η απλούστερη. Εφόσον η κατηγορία action είναι η NOACTION, αποτέλεσμα της σύνθεσης θα είναι είτε το ίδιο το Service αν αποτελείται από δύο συνθετικά είτε συνδυαζόμενο με το γενικό συνθετικό “00000”, σε περίπτωση που αποτελείται από ένα μόνο. Η δεύτερη περίπτωση περιλαμβάνει μια κατηγορία service

που αποτελείται από ένα συνθετικό και μια κατηγορία action. Εφόσον το action αυτό δεν είναι μια από τις δύο ιδιαίτερες περιπτώσεις (AGENTSPECIFIC και TROUBLETECH) η σύνθεση είναι απλή. Οι τρείς τελευταίες περιπτώσεις (3, 4, 5) παρουσιάζουν μια ιδιαιτερότητα που προκύπτει για τις δύο πρώτες από το γεγονός πως τα actions μόνο (AGENTSPECIFIC και TROUBLETECH) αντιστοιχούν σε μια πλήρη προκαθορισμένη κατηγορία ενώ για την τελευταία από το γεγονός πως το service αποτελείται από δύο συνθετικά και η κατηγορία action (σε αντίθεση με την πρώτη περίπτωση) δεν είναι η NOACTION.

Για τη σύνθεση είναι απαραίτητο να προσδιορισθεί κάποιος τρόπος ώστε να μην προκύψουν αποτελέσματα με τρία συνθετικά αφού δεν θα ήταν συγκρίσιμα με τα αποτελέσματα του ενός ταξινομητή. Υπενθυμίζεται πως όλες οι προκαθορισμένες κατηγορίες αποτελούνται από δύο συνθετικά. Τελικά, στις περιπτώσεις 4 και 5 διατηρείται το action ενώ στην περίπτωση 5 διατηρείται το service. Έτσι, το αποτέλεσμα της σύνθεσης εξασφαλίζεται πως ανήκει στις προκαθορισμένες, κάτι που δεν θα συνέβαινε αν σε ορισμένες περιπτώσεις διατηρούσαμε ένα συνθετικό από το service και το συνθέταμε με ένα action. Για παράδειγμα η σύνθεση BALAM_INFO που θα μπορούσε να προκύψει δεν περιλαμβάνεται στις προκαθορισμένες κατηγορίες.

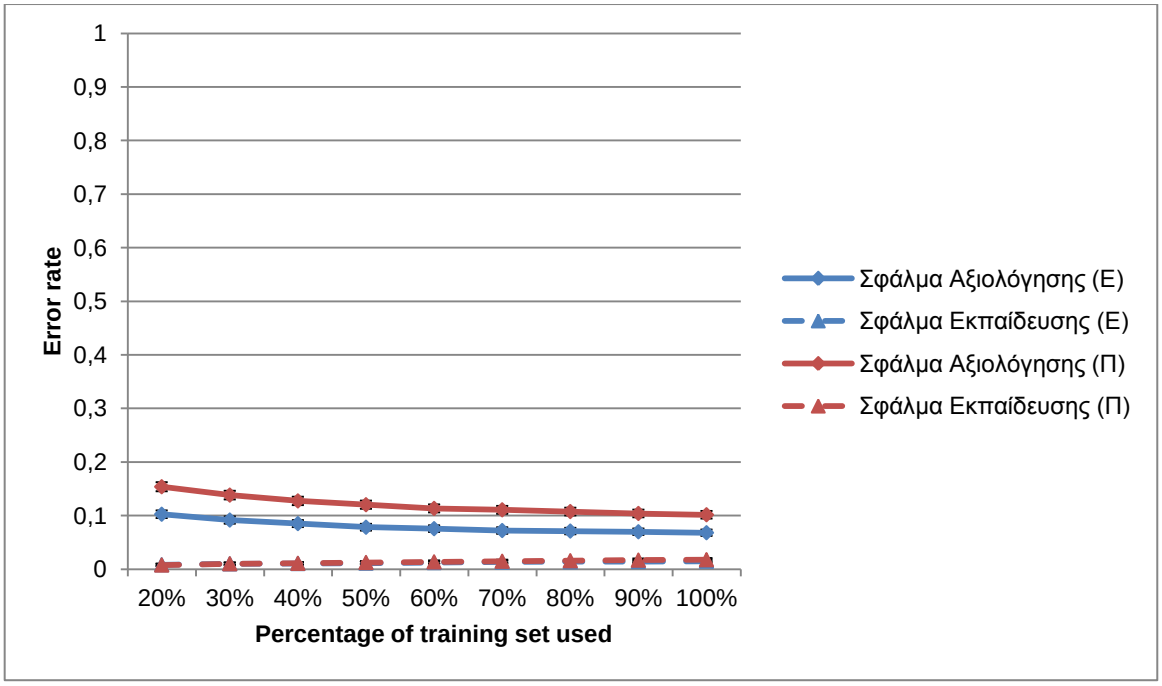
5. ΠΕΙΡΑΜΑΤΑ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1. Καθαρά εκφωνήματα

5.1.1. Προεπιλεγμένες ρυθμίσεις ταξινομητή

Στο συγκεκριμένο πείραμα φαίνεται η διαφορά που προκύπτει χρησιμοποιώντας το ίδιο μοντέλο ταξινομητή αν ο ταξινομητής κληθεί να κατηγοριοποιήσει τα ίδια δεδομένα στις προκαθορισμένες και τις ευρύτερες κατηγορίες.

Στο πρώτο αυτό πείραμα χρησιμοποιήθηκε ο ταξινομητής μεγίστης εντροπίας του Stanford με τις προκαθορισμένες επιλογές του. Στην Εικόνα 13 που ακολουθεί φαίνονται οι καμπύλες μάθησης που προκύπτουν χρησιμοποιώντας τον ίδιο ταξινομητή και τα ίδια δεδομένα σε δύο διαφορετικές περιπτώσεις. Με κόκκινο χρώμα τα αποτελέσματα αφορούν δεδομένα επισημειωμένα με τις προκαθορισμένες κατηγορίες, ενώ με μπλέ χρώμα αφορούν δεδομένα επισημειωμένα με τις ευρύτερες κατηγορίες. Σημειώνεται πως στα παρακάτω πειράματα, τα μεγέθη που ακολουθούνται από το σύμβολο “(Π)” αφορούν τις προκαθορισμένες κατηγορίες, ενώ τα μεγέθη που ακολουθούνται από το σύμβολο “(Ε)” τις ευρύτερες κατηγορίες.



Εικόνα 13. Οι καμπύλες μάθησης που προκύπτουν για micro-averaged σφάλμα

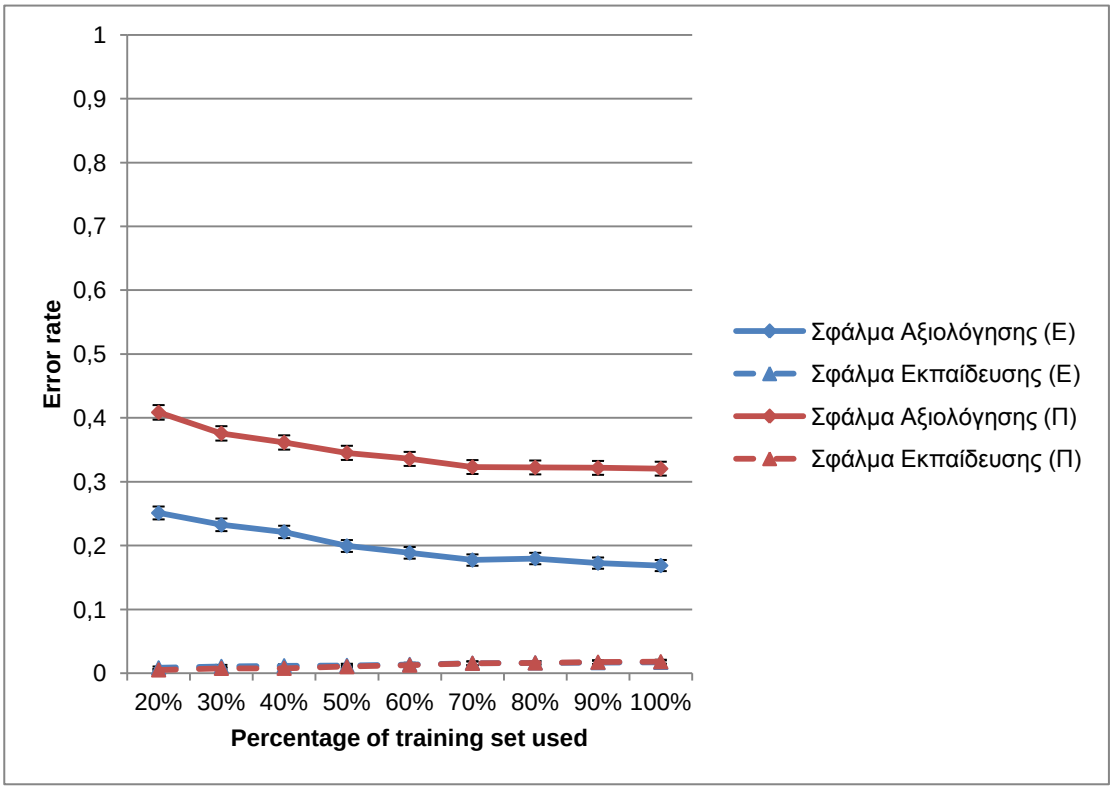
Όσον αφορά στα ίδια τα αποτελέσματα, το πρώτο πράγμα που παρατηρείται είναι πως είναι ικανοποιητικά. Τα σφάλματα εκπαίδευσης σχεδόν ταυτίζονται, με την αντίστοιχη καμπύλη των ευρύτερων κατηγορίων να βρίσκεται κάτω ακριβώς από εκείνη των προκαθορισμένων, και ξεπερνούν ορια κά το 0.02. Αντίθετα στα σφάλματα αξιολόγησης σημειώνεται διαφορά ανάμεσα στις ευρύτερες και τις προκαθορισμένες κατηγορίες ενώ λογικά, εφόσον τα δεδομένα είναι διαφορετικά από αυτά στα οποία εκπαιδεύεται ο ταξινομητής, τα αποτελέσματα είναι χειρότερα φτάνοντας μέχρι και το 0.15.

Η σύγκριση ανάμεσα στις δύο διαφορετικές προσεγγίσεις δείχνει πως το σφάλμα εκπαίδευσης δεν διαφέρει σημαντικά ανάμεσα στις προκαθορισμένες και τις ευρύτερες κατηγορίες. Αυτό σημαίνει πως ο ταξινομητής είναι ικανός να μάθει πώς να κατηγοριοποιεί εξίσου καλά τα δεδομένα πάνω στα οποία εκπαιδεύθηκε τόσο όταν πρόκειται για τις προκαθορισμένες όσο και για τις ευρύτερες κατηγορίες.

Αντίθετα, το σφάλμα αξιολόγησης είναι ελαφρά υψηλότερο για τις προκαθορισμένες κατηγορίες κατά 0.03 περίπου. Όταν δηλαδή ο ταξινομητής κληθεί να κατηγοριοποιήσει νέα δεδομένα, η μεγαλύτερη πολυπλοκότητα των προκαθορισμένων κατηγοριών που προκύπτει από τον μεγαλύτερο αριθμό

κατηγοριών και τον βαθμό εξειδίκευσής τους, φαίνεται πως τον δυσκολεύει ελαφρά περισσότερο.

Οι αντίστοιχες καμπύλες που προκύπτουν αν αντί για το error rate, που όπως έχουμε εξηγήσει είναι ίσο με 1-micro-averaged F1, παρασταθεί το μέγεθος 1-macro-averaged F1 φαίνονται παρακάτω. Στην περίπτωση του macro-averaging, υπολογίζεται ξεχωριστά το σφάλμα (error rate), η ακρίβεια (precision), η ανάκληση (recall) και το F1 κάθε κατηγορίας και κατόπιν υπολογίζεται ο μέσος όρος των τεσσάρων μεγεθών, για όλες τις κατηγορίες. Έτσι όλες οι κατηγορίες έχουν ίσο βάρος, χωρίς να λαμβάνεται υπόψη ότι κάποιες κατηγορίες είναι συχνότερες.



Εικόνα 14. Οι καμπύλες μάθησης που προκύπτουν για macro-averaged σφάλμα

Είναι εμφανές πως τόσο για τις προκαθορισμένες όσο και για τις ευρύτερες κατηγορίες, το σφάλμα αξιολόγησης έχει αυξηθεί σημαντικά γεγονός που οφείλεται στην ιδιομορφία του macro-averaged σφάλματος. Η απόδοση του ίδιου βάρους σε κατηγορίες που είναι σπάνιες και άρα δυσκολεύουν τον ταξινομητή επιφέρει αυτό το αποτέλεσμα. Το σφάλμα εκπαίδευσης δεν έχει μεταβληθεί, με την αντίστοιχη καμπύλη των προκαθορισμένων κατηγοριών να κρύβει εκείνη των ευρύτερων. Ο

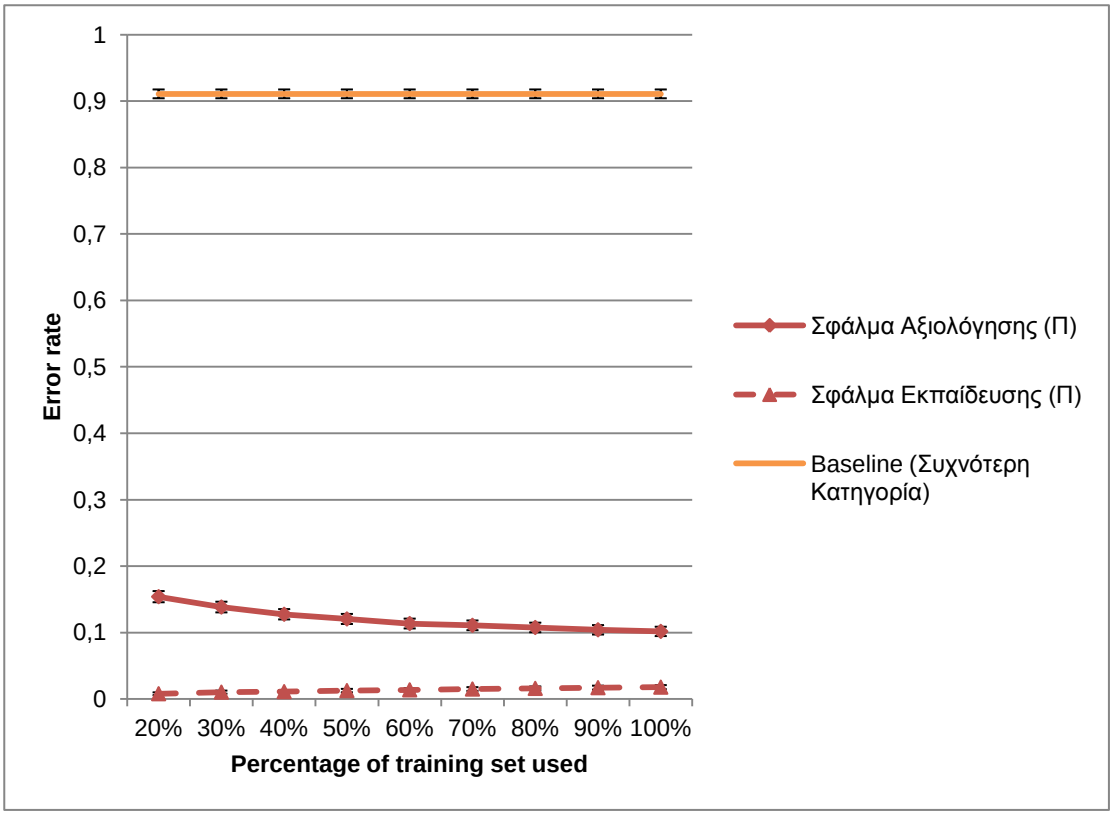
ταξινομητής και πάλι φαίνεται πως πετυχαίνει πολύ καλά αποτελέσματα όταν καλείται να ταξινομήσει τα δεδομένα που χρησιμοποιήθηκαν και για την εκπαίδευσή του.

Τα πειράματα αυτά με τις ευρύτερες κατηγορίες δεν έχουν πρακτικό χαρακτήρα, καθώς μας ενδιαφέρει η κατηγοριοποίηση στις προκαθορισμένες κατηγορίες. Ωστόσο, κρίθηκε σκόπιμο να πραγματοποιηθούν γιατί θα μπορούσαν να χρησιμοποιηθούν σαν ένα άνω όριο για τον ταξινομητή, εφόσον το μοντέλο είναι αρκετά απλούστερο (λιγότερες από μισές κατηγορίες).

5.1.2. Baseline (κατάταξη στην συχνότερη κατηγορία)

Η επόμενη Εικόνα 15 αντιπαραβάλλει τα ίδια αποτελέσματα (για τις προκαθορισμένες κατηγορίες) που σχολιάσθηκαν και προηγουμένως με ένα baseline.

Για να είναι σαφές και ένα κάτω όριο, αποφασίσθηκε, όπως είναι η συνήθης πρακτική σε τέτοιου είδους πειράματα, να εξετασθούν τα αποτελέσματα που θα προέκυπταν αν ο ταξινομητής κατηγοριοποιούσε όλα τα αποτελέσματα στην πλέον συχνά εμφανιζόμενη κατηγορία του συνόλου δεδομένων. Η κατηγορία αυτή είναι η AGENT_00000 με 6.727 εκφωνήματα.



Εικόνα 15. Baseline - κατάταξη στη συχνότερη κατηγορία

Όπως φαίνεται τα αποτελέσματα της baseline μεθόδου είναι πολύ άσχημα. Το σφάλμα πλησιάζει το 0,91, γεγονός απόλυτα λογικό αφού τα 6.727 εκφωνήματα που ανήκουν στην κατηγορία AGENT_00000 αποτελούν περίπου το 1/10 των δεδομένων που χρησιμοποιούνται για να εκπαιδεύσουν τον ταξινομητή.

5.1.3. Παράμετρος Σίγμα

Η δεύτερη σειρά πειραμάτων που πραγματοποιήθηκε αφορούσε την ρύθμιση της παραμέτρου σίγμα του ταξινομητή, η οποία αποτελεί παράμετρο που ελέγχει το κατά πόσον ο ταξινομητής επιχειρεί να αποφύγει ή όχι την υπερεφαρμογή κατά το στάδιο της εκπαίδευσης [STNLP]. Πραγματοποιήθηκαν πειράματα με τις τιμές σ = 3, 5, 30 και 300. Σε κάθε περίπτωση, οι διαφορές στα αποτελέσματα, τόσο στο micro όσο και στο macro σφάλμα, ήταν αμελητέες σε βαθμό που δεν ήταν ορατές στα αντίστοιχα διαγράμματα, οπότε τα τελευταία παραλείφθηκαν.

5.1.4. Ν-γράμματα

Μια ακόμη διαφορετική προσέγγιση που χρησιμοποιήθηκε στα πειράματα ήταν η χρήση n-γραμμάτων (ακολουθιών n λέξεων) σαν features του ταξινομητή. Το σκεπτικό ήταν πως με περισσότερα features ο ταξινομητής θα μπορούσε να λάβει καλύτερες αποφάσεις, κάτι που όμως δεν φάνηκε στην πράξη. Πραγματοποιήθηκαν πειράματα χρησιμοποιώντας μονογράμματα, συνδυασμό μονογραμμάτων και διγραμμάτων και συνδυασμό μονογραμμάτων, διγραμμάτων και τριγραμμάτων.

Αν και προέκυψαν αρκετά περισσότερα features, η μεταβολή των αποτελεσμάτων ήταν μηδαμινή, πιθανότατα γιατί τα διγράμματα, τα τριγράμματα και τα τετραγράμματα εμφανίζονται πολύ σπάνια σε εκφωνήματα. Έτσι, τα αντίστοιχα διαγράμματα παραλείπονται.

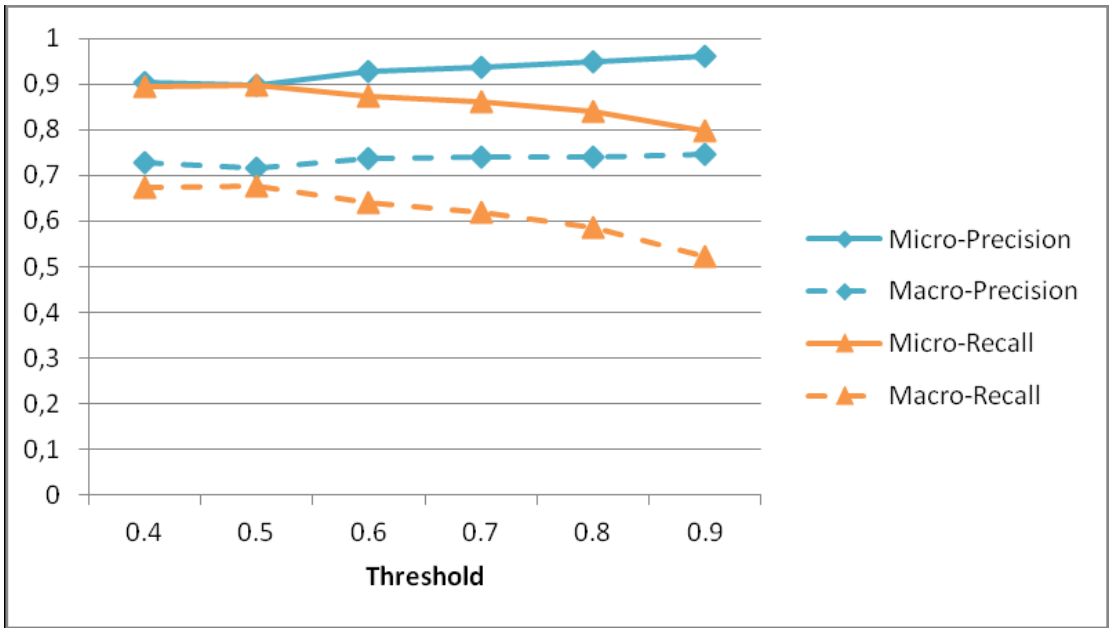
5.1.5. Κατώφλι κατάταξης

Μια ενδιαφέρουσα ιδιότητα του ταξινομητή μεγίστης εντροπίας που χρησιμοποιήθηκε είναι πως το αποτέλεσμα της ταξινόμησης για κάθε δεδομένο (εκφώνημα, στην περίπτωσή μας) δεν είναι μόνο μια κατηγορία, αλλά μια κατανομή πιθανοτήτων, που δείχνει πόσο πιθανό θεωρεί ο ταξινομητής το δεδομένο να ανήκει σε κάθε μία κατηγορία. Το δεδομένο κατατάσσεται στην κατηγορία που ο ταξινομητής θεωρεί πιθανότερη, αλλά η επιστρεφόμενη κατανομή πιθανοτήτων προσφέρει χρήσιμα στοιχεία για τη βεβαιότητα του ταξινομητή.

Πραγματοποιήθηκε σειρά πειραμάτων χαρακτηριστικό των οποίων ήταν η τοποθέτηση κατωφλίου στις πιθανότητες που επιστρέφει ο ταξινομητής. Έτσι, εφόσον για ένα εκφώνημα η πιθανότητα της πιθανότερης (κατά τον ταξινομητή) κατηγορίας δεν ξεπερνούσε το κατώφλι, το εκφώνημα δεν κατατασσόταν σε καμία κατηγορία. Τα πειράματα αυτά δείχνουν τον βαθμό στον οποίο τα αποτελέσματα θα βελτιώνονταν χωρίς τις αποφάσεις που λαμβάνονται με χαμηλή πιθανότητα, καθώς και το πλήθος τον εκφωνημάτων που δυσκολεύουν τον ταξινομητή. Έτσι προσδιορίζεται εάν θα είχε νοήμα να ληφθεί διαφορετική μέριμνα για τα εκφωνήματα που δυσκολεύουν το

ταξινομητή, όπως για παράδειγμα χρήση άλλου ταξινομητή, μη αυτοματοποιημένη ταξινόμηση (παρέμβαση ανθρώπου-τηλεφωνητή) κλπ.

Στην Εικόνα 16 φαίνεται η μεταβολή των μεγεθών micro και macro precision και recall ανάλογα με τις διαφορετικές στάθμες του κατωφλίου.



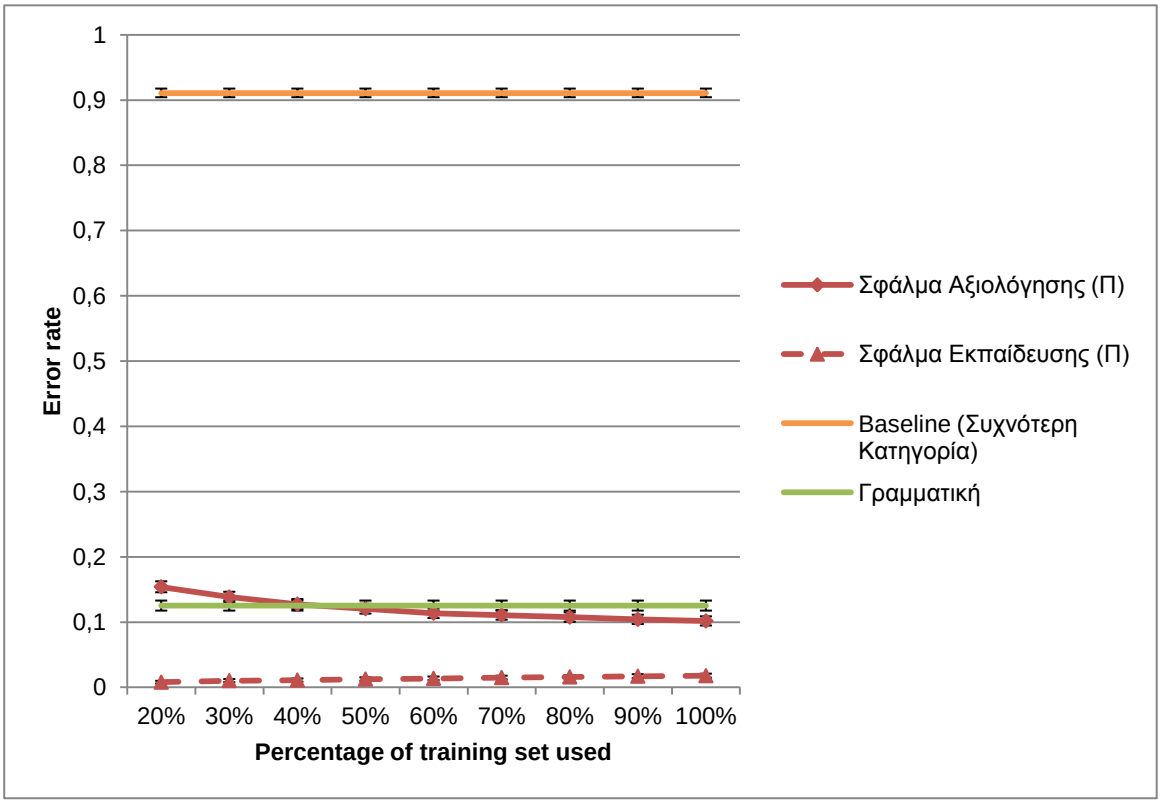
Εικόνα 16. Μεταβολή των micro precision και recall συναρτήσει του κατωφλίου

Όπως αναμενόταν, όσο αυξάνεται το κατώφλι η ακρίβεια (precision) των αποτελεσμάτων αυξάνεται ενώ η ανάκληση (recall) μειώνεται. Επίσης αναμενόμενο είναι το γεγονός πως τα micro αποτελέσματα είναι υψηλότερα από τα macro, λόγω της άνισης κατανομής των εκφωνημάτων στις κατηγορίες.

5.1.6. Σύγκριση του ταξινομητή με το προϋπάρχον σύστημα

Στο διάγραμμα που ακολουθεί επαναλαμβάνονται τα αποτελέσματα δύο πειραμάτων που έχουν ήδη περιγραφεί, όπου χρησιμοποιείται η baseline μέθοδος και ο ταξινομητής μεγίστης εντροπίας, προσθέτοντας τώρα τα αποτελέσματα του προϋπάρχοντος συστήματος που χρησιμοποιείται στην εταιρεία κινητής τηλεφωνίας. Υπενθυμίζεται ότι το τελευταίο βασίζεται σε μια χειρωνακτικά κατασκευασμένη γραμματική.

Είναι ξεκάθαρο πως ο ταξινομητής και η γραμματική επιτυγχάνουν πολύ καλύτερα αποτελέσματα από τη μέθοδο baseline. Ο ταξινομητής εμφανίζει ελάχιστα καλύτερα αποτελέσματα από το προϋπάρχον σύστημα, με την διαφορά όμως να μην είναι σημαντική. Τα error bars αντιστοιχούν σε διαστήματα εμπιστοσύνης 95%. Παρόλα αυτά η χρήση του ταξινομητή εμφανίζει σημαντικά πλεονεκτήματα, καθώς απαιτείται αρκετά λιγότερη δουλειά και χρόνος για να τεθεί σε λειτουργία συγκριτικά με τη χειρωνακτική κατασκευή μιας γραμματικής. Για τον ίδιο λόγο ο ταξινομητής είναι αρκετά ευκολότερο να προσαρμοστεί σε νέα δεδομένα.

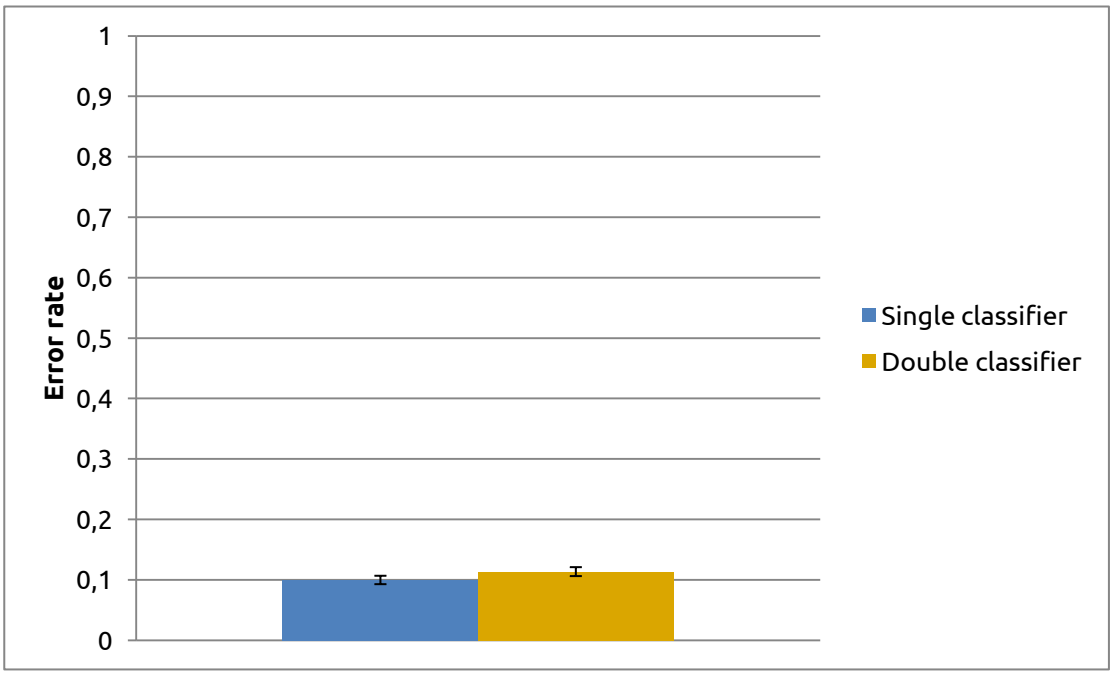


Εικόνα 17. Σύγκριση της γραμματικής με τον ταξινομητή και το baseline

Σημαντικό είναι επίσης το γεγονός πως τα συγκεκριμένα αποτελέσματα αφορούν τα καθαρά εκφωνήματα. Τα θορυβώδη εκφωνήματα αναμένεται να παρουσιάζουν διαφορετικά αποτελέσματα.

5.1.7. Χρήση δύο ταξινομητών έναντι ενός

Τέλος, χρησιμοποιώντας τα καθαρά εκφωνήματα εξετάσθηκε η χρήση δύο ταξινομητών έναντι ενός, όπως περιγράφηκε στο Κεφάλαιο 4.2.4. Τα αποτελέσματα φαίνονται στην Εικόνα 18.



Εικόνα 18. Σύγκριση δύο ταξινομητών έναντι ενός

Όπως φαίνεται, η χρήση ενός ταξινομητή πετυχαίνει ελαφρώς καλύτερα αποτελέσματα, με την διαφορά όμως να είναι αμελητέα. Μάλιστα, αν από τα αποτελέσματα του διπλού ταξινομητή δεν ληφθούν υπόψη οι απαντήσεις εκείνες που αντιστοιχούν σε μη υπάρχουσες κατηγορίες, η διαφορά μειώνεται ακόμη περισσότερο.

5.2. Θορυβώδη εκφωνήματα

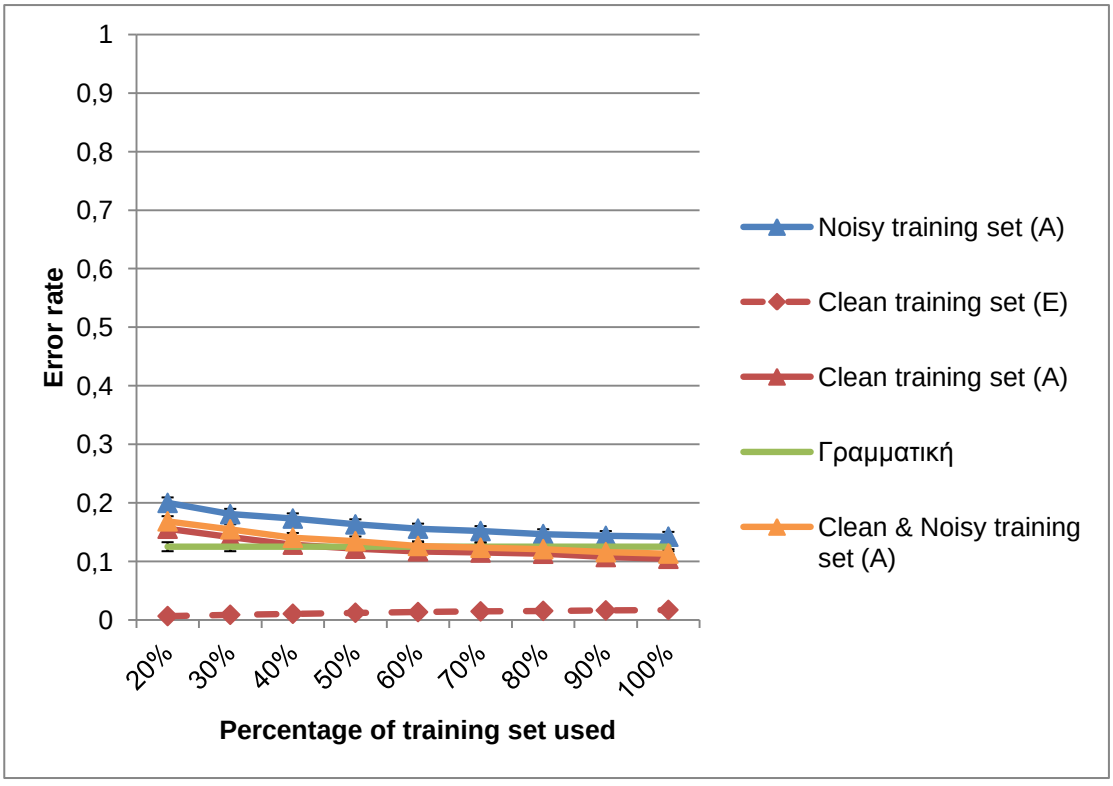
Η ύπαρξη αντίστοιχων εκφωνημάτων σε καθαρή και θορυβώδη μορφή έδωσε την δυνατότητα πραγματοποίησης διαφόρων ενδιαφερόντων δοκιμών εκπαίδευσης και αξιολόγησης του ταξινομητή. Συγκεκριμένα οι δοκιμές που πραγματοποιήθηκαν

φαίνονται συνοπτικά στον Πίνακα που ακολουθεί, ενώ τα αποτελέσματα παρουσιάζονται αναλυτικά στα επόμενα διαγράμματα.

	ΕΚΠΑΙΔΕΥΣΗ	ΑΞΙΟΛΟΓΗΣΗ
1	Καθαρά	Καθαρά
2	Θορυβώδη	Καθαρά
3	Καθαρά	Θορυβώδη
4	Θορυβώδη	Θορυβώδη
5	Καθαρά και Θορυβώδη	Θορυβώδη

Πίνακας 10. Οι διάφορες δοκιμές που πραγματοποιήθηκαν

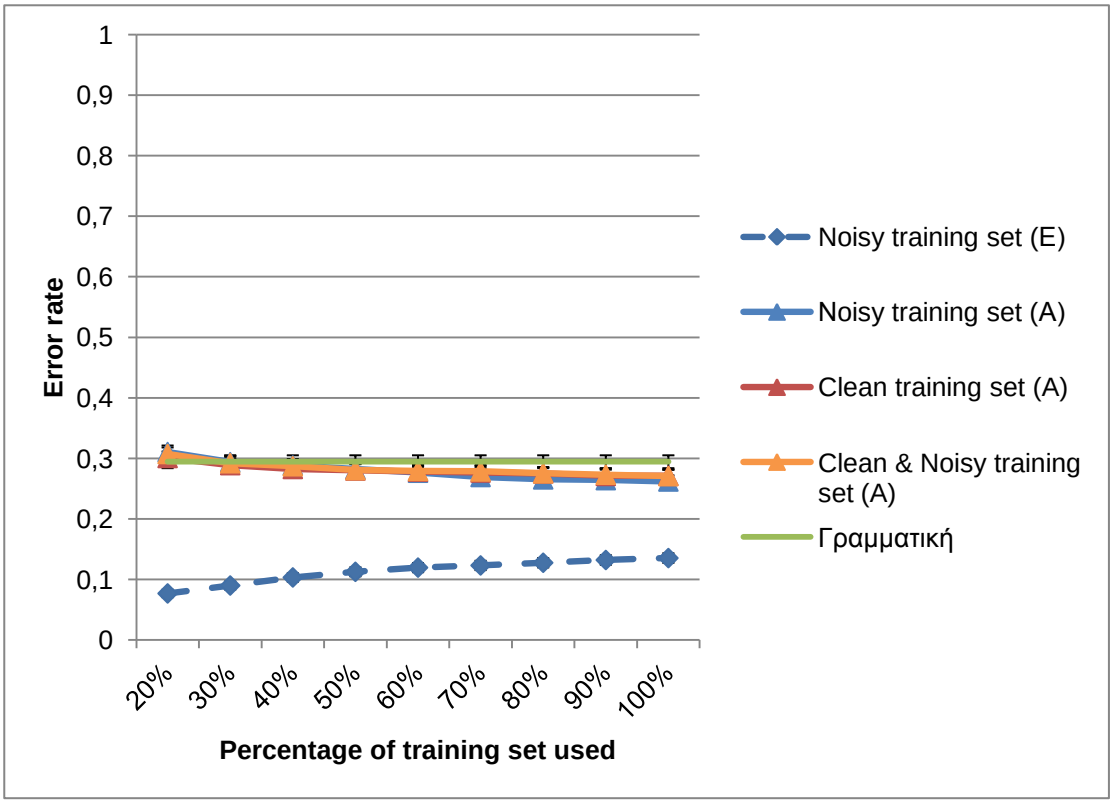
Στο επόμενο διάγραμμα φαίνονται τα αποτελέσματα που προέκυψαν από τους συνδυασμούς 1 και 2 του παραπάνω Πίνακα μαζί με τα αποτελέσματα που προέκυψαν από την γραμματική που χρησιμοποιεί το υπάρχον σύστημα. Τα γράμματα "A" και "E" στις παρενθέσεις ορίζουν αν η αντίστοιχη καμπύλη αποτυπώνει το σφάλμα αξιολόγησης (A) ή το σφάλμα εκπαίδευσης (E).



Εικόνα 19. Αποτελέσματα στα καθαρά εκφωνήματα εκπαιδεύοντας με διαφορετικού τύπου δεδομένα

Όπως και πρίν, ο ταξινομητής που έχει εκπαιδευθεί χρησιμοποιώντας τα καθαρά εκφωνήματα παρουσιάζει ελαφρώς καλύτερα αποτελέσματα από την γραμματική. Αντίθετα, ο ταξινομητής που εκπαιδεύθηκε χρησιμοποιώντας τα θορυβώδη εκφωνήματα πετυχαίνει ελαφρώς χειρότερα αποτελέσματα.

Σε αντιδιαστολή με το παραπάνω, το επόμενο διάγραμμα παρουσιάζει τα αποτελέσματα που προκύπτουν από τους συνδυασμούς 3, 4 και 5 του Πίνακα 10 και τα αντίστοιχα αποτελέσματα της γραμματικής. Είναι δηλαδή μια επισκόπηση των αποτελεσμάτων των διαφόρων ταξινομητών αξιολογώντας τους στα θορυβώδη εκφωνήματα.



Εικόνα 20. Αποτελέσματα στα θορυβώδη εκπαιδεύοντας με διαφορετικού τύπου δεδομένα

Είναι εμφανές πως τόσο η απόδοση των ταξινομητών όσο και της γραμματικής έχουν πέσει αισθητά συγκριτικά με τα καθαρά εκφωνήματα. Η γραμματική αυτή τη φορά πετυχαίνει τα χειρότερα αποτελέσματα χωρίς πάντως η διαφορά από τους υπόλοιπους συνδυασμούς να είναι σημαντική. Τα αποτελέσματα των ταξινομητών σχεδόν ταυτίζονται.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ

Το έναυσμα για την παρούσα διπλωματική εργασία ήταν τα σημαντικά περιθώρια βελτίωσης που παρατηρούνται σε τηλεφωνικά κέντρα, τόσο ως προς την ποιότητα εξυπηρέτησης, όσο και ως προς την αποδοτικότητα. Πιο συγκεκριμένα, εξετάσθηκε η δυνατότητα αντικατάστασης ενός προϋπάρχοντος συστήματος δρομολόγησης κλήσεων, που βασίζεται σε μια χειρωνακτικά κατασκευασμένη γραμματική, από έναν ταξινομητή μεγίστης εντροπίας.

Για τον σκοπό αυτό, πραγματοποιήθηκε επισκόπηση της έρευνας στον τομέα της αυτοματοποιημένης δρομολόγησης κλήσεων και καταγράφηκαν εναλλακτικές λύσεις που έχουν παρουσιαστεί στην βιβλιογραφία για το συγκεκριμένο ζήτημα. Αρχικά, μελετήθηκαν τα δεδομένα και εξήχθησαν ενδιαφέροντα στατιστικά στοιχεία. Στη συνέχεια πραγματοποιήθηκαν διαφόρων τύπων πειράματα, στα οποία μελετήθηκαν οι καμπύλες μάθησης του ταξινομητή μεγίστης εντροπίας. Επιπροσθέτως, εξετάσθηκαν εναλλακτικές τεχνικές, με σκοπό την περαιτέρω βελτίωση των αποτελεσμάτων, που περιελάμβαναν την επισημείωση των δεδομένων με διαφορετικές κατηγορίες και την εξέταση της χρήσης δύο ταξινομητών. Τα πειράματα έδειξαν ότι ο ταξινομητής μπορεί να επιτύχει εξίσου καλά ή και ελαφρώς καλύτερα αποτελέσματα από το προϋπάρχον σύστημα, που βασίζεται στη χειρωνακτικά κατασκευασμένη γραμματική, επομένως προτείνεται η αξιοποίηση του ταξινομητή.

Ενδεχόμενες μελλοντικές κατεθύνσεις για την περαιτέρω αξιοποίηση των δεδομένων είναι οι δοκιμές με διαφορετικούς αλγόριθμους μάθησης και διαφορετικές ιδιότητες (π.χ. tf-idf), αλλά και η στενότερη σύνδεση του συστήματος δρομολόγησης των (μεταγεγραμμένων) εκφωνημάτων με το σύστημα αναγνώρισης φωνής. Για παράδειγμα, θα μπορούσε ενδεχομένως το σύστημα δρομολόγησης να χρησιμοποιεί και ιδιότητες (features) που θα αντανακλούσαν τη βεβαιότητα αναγνώρισης (από το σύστημα αναγνώρισης φωνής) της κάθε λέξης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[BOYE] Boye, J., Wirén, M. (2007). Multi-slot semantics for natural-language call routing systems. In Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 68-75), Rochester, NY, USA.

[BUSE] Busemann, S., Schmeier, S., Arens, R. G. (2000). Message classification in the call center. In Proceedings of the 6th conference on Applied Natural Language Processing (pp. 158-165), Seattle, Washington, USA.

[CARR] Chu-Carroll, J., Carpenter, B. (1999). Vector-based natural language call routing. Computational Linguistics, 25(3), 361-388.

[CHEL1] Chelba, C., Mahajan, M., Acero, A. (2003). Speech utterance classification. In Proceedings of Acoustics, Speech, and Signal Processing (Vol. 1, pp. 280-283), Hong Kong, Hong Kong.

[CHEL2] Chelba, C., Acero, A. (2003). Discriminative training of n-gram classifiers for speech and text routing. In Proceedings of the European Conference on Speech Communication and Technology (Vol. 4, pp. 2777-2780), Geneva, Switzerland.

[CHUC] Chu-Carroll, J., Carpenter, B. (1998). Dialogue management in vector-based call routing. In Proceedings of the 17th International Conference on Computational linguistics (Vol. 1, pp. 256-262), Montreal, Quebec, Canada.

[COX] Cox, S., Shahshahani, B. (2001). A comparison of some different techniques for vector based call-routing. In Proceedings of the European Conference on Speech Communication and Technology (Vol. 4, pp. 2337-2340), Aalborg, Denmark.

[DIFA] Di Fabbrizio, G., Tur, G., Hakkani-Tür, D. (2004). Bootstrapping spoken dialog systems with data reuse. In Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue (pp. 72-80), Cambridge, Massachussetts, USA.

[FERN] Fernández, R., Ginzburg, J., Lappin, S. (2007). Classifying non-sentential utterances in dialogue: A machine learning approach. Computational Linguistics, 33(3), 397-427.

[GLASS] Glass, J., Polifroni, J., Seneff, S., Zue, V. (2000). Data collection and performance evaluation of spoken dialogue systems: The MIT experience. In Proceedings of the 6th International Conference on Spoken Language Processing (Vol. 4, pp. 1-4), Beijing, China.

[GORIN] Gorin, A. L., Riccardi, G., Wright, J. H. (1997). How may I help you?. Speech Communication, 23(1-2), 113-127.

[ΙΑ1] Ανδρουτσόπουλος, Ι. (2011-2012). Διαφάνειες μαθήματος Γλωσσική Τεχνολογία. 3η Ενότητα: Κατάταξη κειμένων και λέξεων. Τμήμα Πληροφορικής, Οικονομικό Πανεπιστήμιο Αθηνών.

[KUO1] Kuo, H. K. J., Lee, C. H., Zitouni, I., Fosler-Lussier, E., & Ammicht, E. (2002). Discriminative training for call classification and routing. In Proceedings of the 7th International Conference on Spoken Language Processing (pp. 1145-1148), Denver, Colorado, USA.

[KUO2] Kuo, H. K., Lee, C. H. (2001). Simplifying design specification for automatic training of robust natural language call router. In Proceedings of Acoustics, Speech, and Signal Processing (Vol. 1, pp. 609-612), Salt Lake City, Utah, USA.

[LEE] Lee, C.J., Chang, J.S. (2002). An Operator Assisted Call Routing System. In Proceedings of the 16th Pacific Asia Conference on Language, Information and Computation (pp. 271-278), Jeju, Korea.

[MANN] Manning, C.D., Raghavan, P., Schutze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.

[NG1] Ng, A. (2012). CS229 Machine Learning. Advice for applying Machine Learning. Stanford University.

[RICC] Riccardi, G., Gorin, A. L., Ljolje, A., & Riley, M. (1997). A spoken language system for automated call routing. In Proceedings of Acoustics, Speech, and Signal Processing (Vol. 2, pp. 1143-1146), Munich, Germany.

[SCWIKI] http://www-nlp.stanford.edu/wiki/Software/Classifier [Accessed 17/01/13]

[STNLP] http://nlp.stanford.edu/index.shtml [Accessed 17/01/13]

[TURU] Turunen, M., Hakulinen, J., Kainulainen, A. (2006). Evaluation of a spoken dialogue system with usability tests and long-term pilot studies: Similarities and differences. In Proceedings of the 9th International Conference on Spoken Language Processing (pp. 1057-1060), Pittsburgh, Pennsylvania, USA.

[WALK] Walker, M. A., Litman, D. J., Kamm, C. A., & Abella, A. (1997). PARADISE: A framework for evaluating spoken dialogue agents. In Proceedings of the 8th Conference of the European Chapter of the Association for Computational Linguistics (pp. 271-280), Madrid, Spain.

[WANG] Wang, Y. Y., Lee, J., Acero, A. (2006). Speech utterance classification model training without manual transcriptions. In Proceedings of Acoustics, Speech and Signal Processing (Vol. 1, pp. I-I), Toulouse, France.