



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΟΛΟΓΩΝ ΚΑΙ ΑΕΡΟΝΑΥΠΗΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΕΝΕΡΓΕΙΑΣ, ΑΕΡΟΝΑΥΤΙΚΗΣ ΚΑΙ ΠΕΡΙΒΑΛΛΟΝΤΟΣ

ΣΠΟΥΔΑΣΤΙΚΗ ΕΡΓΑΣΙΑ

**Πρόβλεψη Σύστασης Προϊόντων με χρήση Μηχανικής Μάθησης και
Τεχνικών Επεξεργασίας Φυσικής Γλώσσας**

Γκιζελής Χαράλαμπος

A.M. : 246524

Επιβλέπων: Καρακαπιλίδης Νικόλαος (Καθηγητής)

Συνεπιβλέπων: Λαζανάς Αλέξης (Μέλος ΕΔΙΠ)

ΠΑΤΡΑ, [Ιούνιος/2021]

ΠΕΡΙΛΗΨΗ**Πρόβλεψη Σύστασης Προϊόντων με χρήση Μηχανικής Μάθησης και
Τεχνικών Επεξεργασίας Φυσικής Γλώσσας****Γκιζελής Χαράλαμπος**

Η ραγδαία ανάπτυξη της τεχνολογίας και ταυτόχρονα του διαδικτύου, τα τελευταία χρόνια, έχει επιφέρει αξιοσημείωτες αλλαγές στον τρόπο επικοινωνίας, ενημέρωσης και φυσικά στον χώρο της αγοράς προϊόντων. Το ηλεκτρονικό εμπόριο, στις μέρες μας, αποτελεί τον κυρίαρχο τρόπο αγορών, ωστόσο ο μεγάλος όγκος πληροφοριών στις ηλεκτρονικές αγορές, αναδεικνύει την πολυπλοκότητα της διαδικασίας. Το συγκεκριμένο «πρόβλημα» καλούνται να αντιμετωπίσουν τα Συστήματα Συστάσεων(Recommender Systems), εντοπίζοντας προϊόντα ή υπηρεσίες σύμφωνα με τα ενδιαφέροντα κάθε χρήστη-πελάτη.

Στη παρούσα Σπουδαστική Εργασία, προσεγγίζουμε το θέμα των συστάσεων-προτάσεων από μια διαφορετική -της καθιερωμένης- σκοπιά, χρησιμοποιώντας Επεξεργασία Φυσικής Γλώσσας. Παρουσιάζουμε εφαρμογές και αλγόριθμους της Μηχανικής Μάθησης, αναλύουμε μεθόδους και τεχνικές της επεξεργασίας φυσικής γλώσσας και κάνουμε μια ανασκόπηση σχετικά με την χρήση της τελευταίας στον τομέα των συστάσεων. Επιπλέον, εισάγουμε τον όρο της ανάλυσης συναισθημάτων και παρουσιάζουμε τα αποτελέσματα που προκύπτουν όταν συνδυαστεί με την επεξεργασία φυσικής γλώσσας.

Λέξεις-Κλειδιά:**Συστήματα Συστάσεων, Μηχανική Μάθηση, Επεξεργασία Φυσικής Γλώσσας,
Ανάλυση Συναισθήματος**

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας,
Αεροναυτικής και Περιβάλλοντος

ABSTRACT**Predicting Items' Recommendation using Machine Learning and
Natural Language Processing Techniques****Gizelis Charalampos**

The rapid development of technology and at the same time of the internet, in recent years, has brought about remarkable changes in the way of communication, information and of course in the field of product marketing. E-commerce, nowadays, is the dominant way of shopping, however, the large volume of information in e-shopping, highlights the complexity of the process. The specific "problem" is called to deal with the Recommender Systems, identifying products or services according to the interests of each user-client.

In this Thesis, we approach the topic of recommendations-proposals from a different- than the established- point of view, using Natural Language Processing (NLP). We present applications and algorithms of Machine Learning, analyze methods and techniques of natural language processing and make a review on the use of the latter in the field of recommendations. In addition, we introduce the term sentiment analysis and present the results that result when combined with natural language processing.

Key-Words:**Natural Language Processing, Machine Learning, Recommender Systems,
Sentiment Analysis**

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1: Απεικόνιση διαδικασίας Sentiment Analysis.....	25
Εικόνα 2: Διαφορετικές Προσεγγίσεις του Sentiment Analysis.....	26
Εικόνα 3: Έρευνα ELTURN και ΙΕΛΚΑ 2020.....	33
Εικόνα 4: Ηλεκτρονικό εμπόριο, 1ο τρίμηνο 2010 – 2019 στην Ελλάδα.....	34
Εικόνα 5: Αναζήτηση προϊόντος στην Amazon.....	35
Εικόνα 6: Collaborative Filtering Recommender Systems.....	38
Εικόνα 7 : Content-Based Recommender System.....	39
Εικόνα 8: Demographic based Recommender System.....	40
Εικόνα 9: Hybrid Recommender System.....	42

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ.....	3
ABSTRACT	4
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	5
ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ	5
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	5
ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ.....	9
1.1 ΗΛΕΚΤΡΟΝΙΚΟ ΕΜΠΟΡΙΟ.....	9
1.1.1 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ ΗΛΕΚΤΡΟΝΙΚΟΥ ΕΜΠΟΡΙΟΥ.....	9
1.1.2 ΟΡΙΣΜΟΣ.....	9
1.1.3 ΠΛΕΟΝΕΚΤΗΜΑΤΑ-ΜΕΙΟΝΕΚΤΗΜΑΤΑ.....	10
1.1.3.1 ΠΛΕΟΝΕΚΤΗΜΑΤΑ.....	11
1.1.3.2 ΜΕΙΟΝΕΚΤΗΜΑΤΑ.....	12
1.2 ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ.....	14
1.2.1 ΕΦΑΡΜΟΓΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ.....	14
1.2.2.1 ΕΠΙΒΛΕΠΟΜΕΝΗ ΜΑΘΗΣΗ (SUPERVISED LEARNING).....	15
1.2.2.2 ΜΗ ΕΠΙΒΛΕΠΟΜΕΝΗ ΜΑΘΗΣΗ (UNSUPERVISED LEARNING)	15
1.2.2.3 ΗΜΙ-ΕΠΟΠΤΕΥΟΜΕΝΗ ΜΑΘΗΣΗ.....	17
1.2.2.3 ON-LINE ΜΑΘΗΣΗ.....	18
Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας, Αεροναυτικής και Περιβάλλοντος	

1.3 ΕΠΕΞΕΡΓΑΣΙΑ ΦΥΣΙΚΗΣ ΓΛΩΣΣΑΣ(NLP).....	19
1.3.1 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ NLP.....	21
1.3.1.1 ΠΡΩΤΗ ΕΠΟΧΗ (1940-1950)	21
1.3.1.2 ΔΕΥΤΕΡΗ ΕΠΟΧΗ(1957-1970)	22
1.3.1.3 ΤΡΙΤΗ ΕΠΟΧΗ(1970-1993).....	22
1.3.1.4 ΤΕΤΑΡΤΗ ΕΠΟΧΗ(1993-MΕΧΡΙ ΣΗΜΕΡΑ).....	23
1.3.2 ΣΗΜΑΝΤΙΚΕΣ ΕΦΑΡΜΟΓΕΣ ΤΟΥ NLP.....	23
1.4 SENTIMENT ANALYSIS Ή OPINION MINING.....	24
ΚΕΦΑΛΑΙΟ 2: ΣΧΕΤΙΚΗ ΕΡΕΥΝΑ(RELATED WORK).....	27
ΚΕΦΑΛΑΙΟ 3: ΗΛΕΚΤΡΟΝΙΚΕΣ ΑΓΟΡΕΣ ΚΑΙ ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ....	31
3.1 ΗΛΕΚΤΡΟΝΙΚΑ ΚΑΤΑΣΤΗΜΑΤΑ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΟ ΕΜΠΟΡΙΟ.....	32
3.2 Η ΕΤΑΙΡΙΑ AMAZON.....	34
3.3 ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ(RS).....	36
3.4 ΚΑΤΗΓΟΡΙΕΣ ΣΥΣΤΗΜΑΤΩΝ ΣΥΣΤΑΣΗΣ.....	37
3.4.1 COLLABORATIVE FILTERING.....	38
3.4.2 CONTENT-BASED.....	39
3.4.3 DEMOGRAPHIC FILTERING.....	40
3.4.4 KNOWLEDGE BASED RS.....	41

3.4.5 HYBRID RS.....	41
ΚΕΦΑΛΑΙΟ 4: ΒΙΒΛΙΟΓΡΑΦΙΑ.....	45

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1 Ηλεκτρονικό εμπόριο

1.1.1 Ιστορική Αναδρομή Ηλεκτρονικού Εμπορίου

Για να προσδιορίσουμε τις ρίζες του ηλεκτρονικού εμπορίου θα χρειαστεί να πάμε αρκετά πίσω καθώς κατά τον 19^ο αιώνα, ο τηλεγράφος ήταν αυτός που συνέβαλλε στη δημιουργία της νέας αυτής έννοιας. Είναι γεγονός, όσο παράξενο και αν φαίνεται, ότι η χρήση του τηλεγράφου έθεσε τα θεμέλια για να δομηθεί το σύγχρονο εμπόριο με ηλεκτρονικά μέσα. Η ανάπτυξη και διάδοση όμως του σύγχρονου εμπορίου οφείλεται κατά βάση στη διάδοση και εμπορευματοποίηση του διαδικτύου. Το Internet, σε συνδυασμό με τις χαμηλές τιμές σε Hardware, ήταν δύο στοιχεία που έδωσαν κίνητρο τόσο στις μεγάλες όσο και στις μικρές επιχειρήσεις να δοκιμάσουν και στη συνέχεια να υιοθετήσουν έναν νέο τρόπο εμπορικής δραστηριότητας. Το internet και συγκεκριμένα το www (World Wide Web), πήρε τη μορφή που έχει σήμερα στα μέσα του 1990 και έχει συμβάλει στην αλματώδη διάδοση αυτού σε κάθε κοινό. Η πρόσβαση σε πληροφορίες και δεδομένα με απλό και εύκολο τρόπο είναι το χαρακτηριστικό που εκτόξευσε την διάδοσή του. Με τη διάδοση λοιπόν του internet η ανάπτυξη επιχειρηματικών εφαρμογών βασισμένη σε ηλεκτρονικά μέσα ήρθε ως μια φυσική συνέχεια της εξέλιξης του εμπορίου. Η συνεργασία μεταξύ ψηφιακής πληροφόρησης, υπολογιστικών εφαρμογών και διαδικτύου άνοιξε τον δρόμο για το σύγχρονο ηλεκτρονικό εμπόριο.

1.1.2 Ορισμός

Πριν ορίσουμε το ηλεκτρονικό εμπόριο ως έννοια, θα πρέπει να αναφέρουμε τα συστατικά στοιχεία για να υπάρξει αυτό. Τα στοιχεία αυτά είναι 4:

- a) Η ύπαρξη επιχείρησης, η οποία καλείται να εκμεταλλευτεί το ηλεκτρονικό εμπόριο και την εξέλιξη της τεχνολογίας με σκοπό οι συναλλαγές και ροή εργασιών να γίνονται όλο και πιο αυτοματοποιημένες.

- b) Αγαθά και υπηρεσίες. Μια επιχείρηση έχει την δυνατότητα να πουλάει αγαθά με υλική υπόσταση όπως είδη ρουχισμού, αλλά και άυλα όπως το δικαίωμα χρήσης ενός λογισμικού. Στόχος όμως του ηλεκτρονικού εμπορίου είναι να ικανοποιήσει την κοινή επιθυμία προμηθευτών και πελατών, που δεν είναι άλλη από την καλύτερη ποιότητα αγαθών και υπηρεσιών σε συνδυασμό με το μικρότερο δυνατό κόστος τους.
- c) Η απόσταση. Το ηλεκτρονικό εμπόριο δίνει τη δυνατότητα, μέσω του διαδικτύου, να γίνονται συναλλαγές αγαθών και υπηρεσιών σε οποιοδήποτε σημείο του κόσμου.
- d) Επικοινωνία. Για να υπάρξει ηλεκτρονικό εμπόριο απαιτείται η παροχή πληροφοριών, προϊόντων ή υπηρεσιών και πληρωμή μέσω ηλεκτρονικού υπολογιστή.

Σε αυτό το σημείο μπορούμε να παρουσιάσουμε, έναν από τους πολλούς ορισμούς για το ηλεκτρονικό εμπόριο που αποδόθηκε από τον Αρσένη Πασχόπουλο:

<< Ως Ηλεκτρονικό εμπόριο ορίζεται η επικοινωνία και η σύναψη εμπορικών συναλλαγών μεταξύ επιχειρήσεων ή μεταξύ επιχειρήσεων και των πελατών τους, με τη χρήση ηλεκτρονικών μέσων. >>

1.1.3 Πλεονεκτήματα – Μειονεκτήματα

Είναι γεγονός πως σήμερα οι φυσικές αγορές έχουν μια τάση συρρίκνωσης, ενώ οι ηλεκτρονικές αγορές εμφανίζουν μια ραγδαία ανάπτυξη στην εξάπλωσή τους. Όπως συμβαίνει με κάθε νέα τεχνολογία και επιχείρηση, έτσι και το ηλεκτρονικό εμπόριο προκαλεί τεράστιες αντιδράσεις. Από την μία πλευρά, έχει τους φανατικούς και ένθερμους υποστηρικτές του, οι οποίοι το χρησιμοποιούν και τους είναι αναγκαίο, ενώ από την άλλη πλευρά έχει τους ένθερμους αντιμάχους του, οι οποίοι δεν θεωρούν ωφέλιμοι την χρήση του ηλεκτρονικού εμπορίου και απορρίπτουν την εφαρμογή του. Παρακάτω θα αναλύσουμε τα πλεονεκτήματα και τα μειονεκτήματα που εμφανίζει το ηλεκτρονικό εμπόριο από την οπτική του καταναλωτή αλλά και από αυτή της επιχείρησης.

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας, Αεροναυτικής και Περιβάλλοντος

1.1.3.1 Πλεονεκτήματα

A) Για καταναλωτές

- 1) Τα ηλεκτρονικά καταστήματα παραμένουν ανοιχτά 24 ώρες το 24ωρο, με αποτέλεσμα ο καταναλωτής να μπορεί να τα “επισκεφτεί” οποιαδήποτε ώρα και μέρα, χωρίς εξαίρεση.
- 2) Παροχή σύγκρισης τιμών. Το ηλεκτρονικό κατάστημα δίνει την δυνατότητα στον χρήστη-καταναλωτή να “επισκεφθεί” πολλά καταστήματα, με αποτέλεσμα να βρει τις καλύτερες τιμές
- 3) Απεριόριστα καταστήματα σε ελάχιστο χρόνο. Το internet δεν έχει γεωγραφικά όρια, έτσι ο κάθε υποψήφιος καταναλωτής μπορεί να επισκεφθεί περισσότερα καταστήματα σε ελάχιστο χρόνο.
- 4) Η ταχύτερη εντόπιση του προϊόντος. Ο καταναλωτής με ένα απλό search στο αντίστοιχο πεδίο, μπορεί να βρει το αντικείμενο που τον ενδιαφέρει.
- 5) Συστήματα Σύστασης. Πλέον, σχεδόν σε όλα τα ηλεκτρονικά καταστήματα, υπάρχει σύστημα που καταγράφει τις ενέργειες που κάνει ο κάθε πελάτης κατά την επίσκεψή του στο κατάστημα. Αυτό έχει ως συνέπεια, κάθε φορά που επισκέπτεται ο καταναλωτής το συγκεκριμένο κατάστημα, να του γίνονται προτάσεις, βασισμένες σε όλες τις προηγούμενες ενέργειές του(κριτική προϊόντος, αγορά, κλπ)

B) Για επιχειρήσεις

- 1) Το ηλεκτρονικό εμπόριο προσφέρει μεγάλη ακρίβεια και ταχύτητα. Τα δύο αυτά στοιχεία δίνουν τη δυνατότητα στις επιχειρήσεις να ανταλλάσσουν γρηγορότερα πληροφορίες και να μειώσουν το κόστος μιας και πλέον οι επιχειρηματικές διαδικασίες γίνονται αυτοματοποιημένα.

2) Ένα άλλο πλεονέκτημα για τις επιχειρήσεις είναι ότι αυτές δεν είναι υποχρεωμένες να κάνουν φυσική παράδοση κάποιων προϊόντων, όπως λογισμικό ή μουσική, αρκεί ο πελάτης να είναι συνεπής στην καταβολή των χρημάτων(με ηλεκτρονικό τρόπο).

3) Τα συστήματα συστάσεων που αναφέρθηκαν προηγουμένως ως πλεονέκτημα των καταναλωτών, μπορούν να εμφανισθούν και σε αυτή τη λίστα, καθώς η κάθε επιχείρηση μπορεί να προσαρμοστεί στις ανάγκες των καταναλωτών και να προσφέρει όλο και περισσότερες υπηρεσίες σε αυτούς.

4) Σημαντικό είναι επίσης πως οι επιχειρήσεις μειώνουν αρκετά τα έξοδά τους με την ανάπτυξη ενός e-shop. Αυτό συμβαίνει διότι μειώνονται δραστικά τα έξοδα σε εγκαταστάσεις και προσωπικό εξυπηρέτησης, τα οποία θα ήταν αναγκαία αν ένα κατάστημα είχε φυσική υπόσταση.

1.1.3.2 Μειονεκτήματα

A) Για καταναλωτές

- 1) Ψυχολογικοί Λόγοι. Πολλές φορές τα ψώνια ή ‘‘μια βόλτα στα μαγαζιά’’ μπορεί να είναι αρκετά ελαφρυντικό και ανακουφιστικό για την ψυχολογία κάποιου(shopping therapy). Η έλλειψη προσωπικής επαφής μας κάνει να αποστασιοποιούμαστε από την κοινωνική ζωή με αποτέλεσμα να νιώθουμε άσχημα.
- 2) Το ηλεκτρονικό εμπόριο δεν επιτρέπει την δοκιμή ενός προϊόντος πριν την αγορά. Το γεγονός ότι δεν μπορούμε να αγγίξουμε ή να δοκιμάσουμε ένα προϊόν μας κάνει διστακτικούς ως προς την αγορά του. Χαρακτηριστικό παράδειγμα είναι τα παπούτσια ή τα ρούχα.
- 3) Καθυστέρηση των αγαθών. Σε αντίθεση με τα φυσικά καταστήματα όπου υπάρχει άμεση παραλαβή του προϊόντος, στα ηλεκτρονικά καταστήματα υπάρχει αναμονή τουλάχιστον

μία ημέρα. Έτσι, αν κάποιος θέλει να αποκτήσει κάτι άμεσα και να το έχει στα χέρια του, δεν μπορεί να το αποκτήσει από ηλεκτρονικό κατάστημα.

- 4) Το ηλεκτρονικό εμπόριο δεν απευθύνεται σε όλους, αλλά μόνο σε αυτούς που ξέρουν να χρησιμοποιούν το διαδίκτυο και γνωρίζουν τους κινδύνους που ελλοχεύουν. Για την ηλεκτρονική αγορά, απαιτούνται προπληρωμένες ή πιστωτικές κάρτες που αφενός κάποιοι δεν διαθέτουν και αφετέρου υπάρχουν κίνδυνοι κατά την διάθεση τόσο προσωπικών στοιχείων στο διαδίκτυο.

B) Για τις επιχειρήσεις

- 1) Το σημαντικότερο μειονέκτημα για τις επιχειρήσεις είναι το γεγονός ότι οι καταναλωτές δεν εμπιστεύονται ιδιαίτερα το διαδίκτυο για τις συναλλαγές τους και φοβούνται να στείλουν τα στοιχεία των καρτών τους, πράγμα αναγκαίο για μια ηλεκτρονική αγορά.
- 2) Τα περισσότερα καταστήματα εμπορεύονται μικρά αντικείμενα τα οποία είναι εύκολα στην μεταφορά και το κόστος μικρό, ενώ τα καταστήματα που εμπορεύονται βαριά και μεγάλα σε όγκο αντικείμενα είναι ελάχιστα.
- 3) Μειονέκτημα αποτελεί η συνεχώς αναπτυσσόμενη τεχνολογία καθώς το κόστος για να δημιουργήσει μια επιχείρηση ένα απλό ηλεκτρονικό κατάστημα μεταβάλλεται δραματικά μέρα με τη μέρα. Επίσης, μεγάλο κόστος υπάρχει στην συντήρηση και βελτιστοποίηση του ηλεκτρονικού καταστήματος, ποσά που πρέπει να ξοδεύονται σε τακτά χρονικά διαστήματα.
- 4) Άλλο ένα μειονέκτημα που αντιμετωπίζουν οι επιχειρήσεις που θέλουν να αναπτύξουν ένα ηλεκτρονικό κατάστημα είναι η δυσκολία ενσωμάτωσης των βάσεων δεδομένων που ήδη χρησιμοποιούν για τις συναλλαγές τους στο φυσικό κατάστημα,

με λογισμικό το οποίο θα επιτρέπει την ενοποίηση με το ηλεκτρονικό κατάστημα. Για να γίνει αυτό, αναγκαία είναι η επένδυση κεφαλαίου.

1.2 Μηχανική Μάθηση (Machine Learning)

Η μηχανική μάθηση και η τεχνητή νοημοσύνη ήταν πολύ κοντά σαν έννοιες, αλλά πλέον μπορούμε να πούμε με ασφάλεια, ότι η μηχανική μάθηση είναι μία υποκατηγορία αυτής. Στη μηχανική μάθηση εκπαιδεύουμε μοντέλα τα οποία μαθαίνουν από τα δεδομένα, δηλαδή μελετάει και κατασκευάζει αλγορίθμους που μελετούν τα δεδομένα και κάνουν προβλέψεις βασισμένες σε αυτά. Το 1959, η μηχανική μάθηση ορίζεται από το Arthur Samuel ως : «Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί». Το machine learning έχει ως στόχο τη δημιουργία μοντέλων ικανών να μαθαίνουν, να βελτιώνονται δηλαδή μέσω της αξιοποίησης προηγούμενης γνώσης και εμπειρίας. Η εμπειρία αυτή αναφέρεται στις προηγούμενες πληροφορίες που έχει στη διάθεσή του ο εκπαιδευόμενος και έχει τη μορφή ηλεκτρονικών δεδομένων που συλλέγονται και διατίθενται για ανάλυση. Η μηχανική μάθηση συνίσταται στο σχεδιασμό αποτελεσματικών και ακριβών αλγόριθμων πρόβλεψης και η επιτυχία αυτών συνδέεται εγγενώς με τον όγκο και την ποιότητα των δεδομένων και ως επέκταση με την ανάλυση των δεδομένων και τα στατιστικά στοιχεία που προκύπτουν από αυτά.

1.2.1 Εφαρμογές Μηχανικής Μάθησης

Η μηχανική μάθηση εφαρμόζεται σε πληθώρα επιστημών, οι οποίες θα αναφερθούν παρακάτω. Ωστόσο, είναι μια ‘ζωντανή’ επιστήμη και μέρα με τη μέρα βρίσκει εφαρμογή σε όλο και περισσότερους τομείς. Οι πιο διαδεδομένες εφαρμογές της είναι:

- στην αναγνώριση ομιλίας,
- στην αναγνώριση προσώπων,
- στον εντοπισμό προσώπων,
- στην ιατρική για την διάγνωση ασθενειών,

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας,
Αεροναυτικής και Περιβάλλοντος

- στο διαδίκτυο από αρκετές ιστοσελίδες μουσικής και ταινιών προκειμένου να συστήνονται στους επισκέπτες των συγκεκριμένων ιστοσελίδων μουσική και ταινίες για τις οποίες είναι πιθανό να ενδιαφέρονται,
- από τις μηχανές αναζήτησης του διαδικτύου για την διόρθωση ορθογραφικών και συντακτικών λαθών,
- στα αυτόνομα αυτοκίνητα

1.2.2 Κατηγορίες Μηχανικής Μάθησης

Η μηχανική μάθηση (machine learning) είναι ένας τομέας της επιστήμης των υπολογιστών ο οποίος ασχολείται με την μελέτη και την δημιουργία αλγόριθμων οι οποίοι μπορούν να εκπαιδευτούν από τα υπάρχοντα δεδομένα, που συνήθως προέρχονται από πηγές όπως κάποιο αισθητήρα ή ένα σύνολο δεδομένων και να κάνουν προβλέψεις βασισμένες σε αυτό, ουσιαστικά οι αλγόριθμοι μηχανικής μάθησης βελτιώνονται με την εμπειρία. Ανάλογα με το πρόβλημα επιλέγεται το κατάλληλο κριτήριο που θα εφαρμοστεί, από το οποίο εξαρτάται και η αποτελεσματικότητα του. Οι κατηγορίες μηχανικής μάθησης είναι παρόμοιες και είναι κατά κάποιο τρόπο όμοιες με τον τρόπο που λειτουργεί και μαθαίνει ο άνθρωπος. Αυτά τα σενάρια διαφέρουν στους τύπους δεδομένων κατάρτισης που διαθέτει ο εκπαιδευόμενος, τη σειρά και τη μέθοδο από την οποία τα δεδομένα κατάρτισης λαμβάνονται και τα δεδομένα δοκιμών που χρησιμοποιούνται για την αξιολόγηση του αλγορίθμου μάθησης.

1.2.2.1 Επιβλεπόμενη Μάθηση (Supervised Learning)

Σε αυτή την κατηγορία μάθησης, το σύνολο των δεδομένων αποτελείται από ένα πεπερασμένο αριθμό παραδειγμάτων (instances) και το κάθε παράδειγμα αναπαρίσταται από ένα n – διάστατο διάνυσμα $X_i = (x_{i1}, x_{i2} \dots x_{in})$ του οποίου τα στοιχεία ονομάζονται attributes. Επιπλέον, σε κάθε παράδειγμα X_i αντιστοιχεί μία κλάση (class) Y_i . Το σύνολο των παραδειγμάτων που χρησιμοποιείται για την εκπαίδευση του αλγόριθμου ονομάζεται σύνολο εκπαίδευσης (training set). Ο αλγόριθμος κατασκευάζει ένα μοντέλο (μια συνάρτηση) το οποίο απεικονίζει κάθε παράδειγμα σε μια κλάση. Στόχος του μοντέλου αυτού είναι όταν του δοθεί ως είσοδος

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας,
Αεροναυτικής και Περιβάλλοντος

ένα καινούριο παράδειγμα, για το οποίο δεν γνωρίζει την κλάση του, να το κατηγοριοποιήσει (να προβλέψει σωστά την κλάση του) σωστά. Από εκεί και πέρα, για να εξετάσουμε αν το μοντέλο εκτελεί σωστές προβλέψεις, χρησιμοποιούμε ένα σύνολο ελέγχου (test set), το οποίο αποτελείται από τα παραδείγματα τα οποία δεν αξιοποιήθηκαν στο training set. Το σύνολο ελέγχου δίνεται ως είσοδος στο μοντέλο, το οποίο σε κάθε παράδειγμα X_i του συνόλου ελέγχου αντιστοιχίζει μία κλάση Y_i . Το ποσοστό των σωστά κατηγοριοποιημένων παραδειγμάτων του συνόλου ελέγχου είναι και το ποσοστό επιτυχίας των προβλέψεων του μοντέλου.

Η Επιβλεπόμενη μάθηση χρησιμοποιείται σε προβλήματα όπως:

- Ταξινόμηση (Classification)

Αφορά στην δημιουργία μοντέλων πρόβλεψης διακριτών κατηγοριών. Τα δεδομένα εισόδου θα χωριστούν το ελάχιστο σε δυο κλάσεις και το μοντέλο που θα κατασκευαστεί, θα αντιστοιχίζει τα δεδομένα σε μία ή περισσότερες (multi-label ταξινόμηση) κλάσεις. Για παράδειγμα, στην ταξινόμηση εγγράφων μπορεί να αντιστοιχηθούν τα αντικείμενα σε κατηγορίες όπως η πολιτική, οι επιχειρήσεις, ο αθλητισμός ή ο καιρός, ενώ στην ταξινόμηση εικόνων μπορεί να αντιστοιχηθούν αντικείμενα σε κατηγορίες όπως το τοπίο, το πορτρέτο, ή ζώο.

- Παλινδρόμηση (Regression)

Αφορά την δημιουργία μοντέλων πρόβλεψης αριθμητικών τιμών. Στην Παλινδρόμηση αντί για προβλήματα κατηγοριοποίησης ο αλγόριθμος χρησιμοποιείται για προβλήματα παλινδρόμησης. Ένα παράδειγμα Regression είναι η πρόβλεψη της τιμής του σπιτιού, όταν δίνονται τα χαρακτηριστικά του σπιτιού όπως μέγεθος, τετραγωνικά, τιμή κλπ.

- Κατάταξης (Ranking)

Σχετίζεται με την κατανομή των δεδομένων εισόδου σε κάποιο χώρο. Στην αναζήτηση στο Web, η επιστροφή ιστοσελίδων που σχετίζονται με ένα ερώτημα

αναζήτησης αποτελούν παράδειγμα κατάταξης όπως και συστήματα επεξεργασίας φυσικής γλώσσας.

1.2.2.2 Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning)

Στη συγκεκριμένη κατηγορία μάθησης, ο αλγόριθμος πρέπει να υπολογίσει την δομή των δεδομένων που καταχωρούνται, χωρίς να παρέχεται κάποια εμπειρία στον αλγόριθμο μάθησης. Ο αλγόριθμος ουσιαστικά 'ψάχνει' για ένα μοτίβο στα δεδομένα για να καταφέρει να ορίσει κατηγορίες. Χρησιμοποιείται σε προβλήματα:

- Ομαδοποίησης (Clustering)

Η ομαδοποίηση είναι το έργο της διαίρεσης των δεδομένων σε έναν αριθμό ομάδων, έτσι ώστε τα δεδομένα που βρίσκονται στις ίδιες ομάδες να μοιάζουν περισσότερο μεταξύ τους και αυτά που μοιάζουν λιγότερο σε διαφορετικές ομάδες.

- Μείωση διαστασιμότητας (Dimensionality Reduction)

Η μείωση διαστάσεων αναφέρεται σε τεχνικές για τη μείωση του αριθμού των μεταβλητών εισαγωγής στα δεδομένα εκπαίδευσης και ταυτόχρονα διατήρηση των σημαντικών ιδιοτήτων των αρχικά εισαχθέντων δεδομένων. Λιγότερες διαστάσεις εισόδου συχνά σημαίνουν αντίστοιχα λιγότερες παραμέτρους ή μια απλούστερη δομή στο μοντέλο μηχανικής εκμάθησης, που αναφέρεται ως βαθμοί ελευθερίας. Τα δεδομένα δηλαδή απλοποιούνται και αντιστοιχίζονται σε ένα χώρο λιγότερων διαστάσεων.

1.2.2.3 Ημι-εποπτευόμενη μάθηση (Semi supervised Learning)

Η ημι-εποπτευόμενη μάθηση είναι μια προσέγγιση στη μηχανική μάθηση που συνδυάζει μια μικρή ποσότητα δεδομένων με επισήμανση(labeled) με μια μεγάλη ποσότητα δεδομένων χωρίς σήμανση(unlabeled) κατά τη διάρκεια της εκπαίδευσης. Η ημι-εποπτευόμενη μάθηση εμπίπτει μεταξύ της μη εποπτευόμενης μάθησης (χωρίς

ετικέτες εκπαιδευτικών δεδομένων) και της εποπτευόμενης μάθησης (με μόνο επισημασμένα δεδομένα εκπαίδευσης).

1.2.2.3 On-Line Μάθηση(On-Line Learning)

Σε σύγκριση με τις «παραδοσιακές» λύσεις ML, η On-line μάθηση είναι μια θεμελιωδώς διαφορετική προσέγγιση, η οποία αγκαλιάζει το γεγονός ότι τα μαθησιακά περιβάλλοντα μπορούν να αλλάζουν από στιγμή σε στιγμή. Είναι δύσκολο να γίνει σωστά, αλλά όταν συμβεί αυτό, τα αποτελέσματα που μπορούμε να πετύχουμε με τη on-line μάθηση είναι αξιοσημείωτα.

Σε γενικές γραμμές, χρειαζόμαστε δύο πράγματα για τη μηχανική μάθηση: δεδομένα και έναν κατάλληλο αλγόριθμο μάθησης. Ο αλγόριθμος εκμάθησης εκπαιδεύει τα δεδομένα μας και παράγει ένα ακριβές μοντέλο, που χρησιμοποιείται συνήθως για την πρόβλεψη πάνω σε νέα δεδομένα.

Τα μοντέλα ML, εκτός από ορισμένες εξαιρέσεις, είναι στατικά πράγματα. Είναι ουσιαστικά συλλογές παραμέτρων. Αφού εκπαιδεύσουμε ένα μοντέλο, οι παράμετροι του δεν αλλάζουν.

Τι γίνεται όμως με τα νέα δεδομένα; Εάν απλώς εκπαιδεύσουμε ένα μοντέλο μία φορά και δεν το αγγίζουμε ξανά, χάνουμε τις πληροφορίες που θα μπορούσαν να μας παρέχουν περισσότερα δεδομένα. Αυτό είναι ιδιαίτερα σημαντικό σε περιβάλλοντα όπου οι συμπεριφορές αλλάζουν γρήγορα. Ένα καλό παράδειγμα τέτοιου περιβάλλοντος είναι οι διαδικτυακές αγορές: ένα προϊόν που είναι δημοφιλές σήμερα μπορεί να είναι ξεχασμένο αύριο.

Για να μπορέσουμε να επέμβουμε σε αυτό το πρόβλημα και να δημιουργήσουμε ένα ΑΙ που θα <μαθαίνει> συνεχώς, μπορούμε να προγραμματίσουμε εκπαίδευση στα νέα δεδομένα ανά τακτά χρονικά διαστήματα. Το πρόβλημα εδώ είναι ότι ακόμα και αν εκπαιδεύουμε νέα μοντέλα κάθε εβδομάδα ή ακόμα και κάθε μέρα, εξακολουθούμε να μένουμε πίσω. Το μοντέλο μας δεν είναι ποτέ πλήρως ενημερωμένο με τα τρέχοντα γεγονότα, επειδή είναι εκπαιδευμένο σε παλιά δεδομένα. Στην ιδανική περίπτωση, αυτό που θέλουμε είναι ένα μοντέλο που μπορεί να μάθει από νέα παραδείγματα σε κάτι κοντά στον πραγματικό χρόνο. Όχι μόνο να προβλέπει σε πραγματικό χρόνο, αλλά να μαθαίνει και σε πραγματικό χρόνο.

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας,
Αεροναυτικής και Περιβάλλοντος

Θα ήταν δυνατόν να μάθουμε από νέα παραδείγματα σε κάτι κοντά στον πραγματικό χρόνο; Εκεί βρίσκεται η δυνατότητα της διαδικτυακής μάθησης: τη στιγμή που βλέπουμε ένα νέο παράδειγμα, ας μάθουμε από αυτό όσο πιο γρήγορα μπορούμε. Όσο πιο γρήγορα, τόσο το καλύτερο. Στην πραγματικότητα, επειδή η ταχύτητα υπερβαίνει όλα τα άλλα στη διαδικτυακή μάθηση, συνήθως χρησιμοποιούμε απλούς αλγόριθμους μάθησης και όχι περίπλοκους, όπως τα νευρωνικά δίκτυα.

Συνοψίζοντας, το On-line Learning έχει απεριόριστες δυνατότητες με βασικό του μειονέκτημα τη δυσκολία που εμφανίζει στην εφαρμογή του. Η διαδικτυακή μάθηση είναι ένας ριζικά διαφορετικός τρόπος προσέγγισης της μηχανικής μάθησης. Είναι μια προσέγγιση που αγκαλιάζει την αλλαγή, ανεξάρτητα από το πόσο δραστική αυτή είναι.

1.3 Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing)

Οι φυσικές γλώσσες είναι οι γλώσσες που έχουν εξελιχθεί φυσικά και χρησιμοποιούνται από τα ανθρώπινα όντα για λόγους επικοινωνίας, για παράδειγμα, τα αγγλικά, τα γαλλικά και τα γερμανικά είναι φυσικές γλώσσες. Η επεξεργασία φυσικής γλώσσας ή το nlp (ονομάζεται επίσης υπολογιστική γλωσσολογία) είναι η επιστημονική μελέτη των γλωσσών από την υπολογιστική προοπτική. Η επεξεργασία φυσικής γλώσσας (nlp) είναι ένα πεδίο της επιστήμης των υπολογιστών και της γλωσσολογίας που σχετίζεται με τις αλληλεπιδράσεις μεταξύ υπολογιστών και ανθρώπινων γλωσσών. Τα συστήματα παραγωγής φυσικών γλωσσών μετατρέπουν πληροφορίες από βάσεις δεδομένων υπολογιστών σε επαναλαμβανόμενη ανθρώπινη γλώσσα σε πιο επίσημες αναπαραστάσεις, όπως η ανάλυση δέντρων ή η λογική πρώτης τάξης που είναι ευκολότερο να χειριστούν τα προγράμματα υπολογιστών. πολλά προβλήματα στο nlp ισχύουν τόσο για τη δημιουργία όσο και για την κατανόηση.

Το NLP έχει σημαντική αλληλεπικάλυψη με το πεδίο της υπολογιστικής γλωσσολογίας και συχνά θεωρείται υπο-πεδίο της τεχνητής νοημοσύνης. Ο όρος φυσική γλώσσα χρησιμοποιείται για τη διάκριση ανθρώπινων γλωσσών (όπως ισπανικά ή σουηδικά) από επίσημες ή υπολογιστικές γλώσσες (όπως C ++, Java ή LISP). Αν και το nlp μπορεί να περιλαμβάνει κείμενο και ομιλία, η επεξεργασία της ομιλίας γίνεται συμβατικά σε ξεχωριστό πεδίο

Στο $\eta\mu\epsilon\rho$, αναπτύσσονται οι τεχνικές που στοχεύουν στον υπολογιστή να κατανοεί τις εντολές που δίνονται στη φυσική γλώσσα και να εκτελεί ανάλογα. Προς το παρόν, για να δουλέψουμε με υπολογιστή, η εισαγωγή πρέπει να παρέχεται σε επίσημες γλώσσες. Οι επίσημες γλώσσες είναι εκείνες οι γλώσσες που έχουν αναπτυχθεί ειδικά για να επικοινωνούν με υπολογιστή και κατανοούνται από τη μηχανή (FORTRAN, Pascal, κ.λπ.). Προφανώς, για την επικοινωνία με τον υπολογιστή, απαιτείται η μελέτη αυτών των επίσημων γλωσσών. Η κατανόηση αυτών των γλωσσών είναι δυσκίνητη και απαιτεί πρόσθετες προσπάθειες για την κατανόησή τους. Ως εκ τούτου, περιορίζει τις εφαρμογές τους στον υπολογιστή. Σε σύγκριση με αυτό, η επικοινωνία στη φυσική γλώσσα θα διευκολύνει τη λειτουργία και την επικοινωνία με τον υπολογιστή εύκολα και με φιλικό προς τον χρήστη τρόπο.

Η επεξεργασία φυσικής γλώσσας είναι ένας σημαντικός τομέας της τεχνητής νοημοσύνης, επειδή ένας υπολογιστής θα θεωρείται ευφυής εάν μπορεί να κατανοήσει τις εντολές που δίνονται στη φυσική γλώσσα αντί για C, Fortran ή pascal. Ως εκ τούτου, με την ικανότητα των υπολογιστών να κατανοούν τη φυσική γλώσσα, γίνεται πολύ ευκολότερη η επικοινωνία με τους υπολογιστές. Επίσης, η επεξεργασία της φυσικής γλώσσας μπορεί να εφαρμοστεί ως εργαλείο παραγωγικότητας σε εφαρμογές που κυμαίνονται από τη σύνοψη ειδήσεων έως τη μετάφραση από τη μία γλώσσα στην άλλη. Ωστόσο, η επεξεργασία επιφανειακών επιπέδων φυσικών γλωσσών φαίνεται να είναι εύκολη, η βαθιά επεξεργασία φυσικών γλωσσών, η κατανόηση των σιωπηρών μηνυμάτων και οι προθέσεις ενός ομιλητή είναι εξαιρετικά δύσκολες εργασίες.

Οι παλαιότεροι υπολογιστές ήταν κυρίως επεξεργαστές αριθμών και οι διαθέσιμοι πόροι σε αυτούς δεν ήταν διαφορετικοί από αυτούς που ήταν διαθέσιμοι στην προγραμματιζόμενη αριθμομηχανή πρώτης γενιάς. Οι εξελίξεις στον τομέα της επιστήμης των υπολογιστών από τότε που ξεκίνησαν, μας έδωσαν γλώσσες προγραμματισμού που διευκόλυναν τη συζήτηση για τους γλωσσικούς στόχους, όπως λέξεις και προτάσεις

Για πολλούς, η ικανότητα του υπολογιστή να επεξεργάζεται τη γλώσσα τόσο επιδέξια όσο εμείς θα σηματοδοτεί την άφιξη πραγματικά έξυπνων μηχανών. Η βάση αυτής της πεποίθησης είναι το γεγονός ότι η αποτελεσματική χρήση της γλώσσας είναι συνυφασμένη με τις γενικές γνωστικές μας ικανότητες. Μεταξύ των πρώτων που

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας,
Αεροναυτικής και Περιβάλλοντος

εξέτασαν τις υπολογιστικές συνέπειες αυτής της στενής σύνδεσης ήταν ο Άλαν Τούρινγκ. Στο διάσημο τεστ Turing του πρότεινε μια εμπειρική δοκιμή για να κρίνει την ευφυΐα ενός υπολογιστή με τη μορφή ενός παιχνιδιού που περιλαμβάνει αλληλεπίδραση ανθρώπου-υπολογιστή. Εάν ο υπολογιστής κερδίσει το παιχνίδι, μπορεί να χαρακτηριστεί ως «έξυπνος».

1.3.1 Ιστορική Αναδρομή NLP

Η επεξεργασία της φυσικής γλώσσας(NLP) είναι συγκριτικά μια νέα περιοχή. Αυτή η ενότητα περιγράφει εν συντομία διάφορες αναπτυξιακές φάσεις της επεξεργασίας φυσικής γλώσσας.

1.3.1.1 Πρώτη Εποχή (1940-1950)

Σε αυτήν την περίοδο, η εργασία περιοριζόταν κυρίως σε δύο βασικά παραδείγματα: τον αυτοματισμό και τη χρήση μοντέλων πιθανοτήτων. Ο αυτοματισμός ξεκίνησε το 1950 με το μοντέλο αλγοριθμικού υπολογισμού του Turing. Το έργο του Turing οδήγησε πρώτα στην ανάπτυξη των νευρωνικών του Mc Culloch Pitts, οι οποίοι έκαναν ένα απλοποιημένο μοντέλο νευρώνα ως ένα είδος υπολογιστικού στοιχείου που θα μπορούσε να περιγραφεί με όρους προτατικής λογικής και στη συνέχεια στο έργο του Kleene σχετικά με την πεπερασμένη αυτοματοποίηση και την τακτική έκφραση. Ο Shannon εφάρμοσε πιθανολογικά μοντέλα διακριτών διαδικασιών Markov για αυτοματοποίηση της επεξεργασίας γλώσσας. Αντλώντας την ιδέα της πεπερασμένης διαδικασίας Markov από το έργο του Shannon, ο Chomsky, θεωρούσε για πρώτη φορά τις μηχανές πεπερασμένων καταστάσεων ως έναν τρόπο χαρακτηρισμού μιας γραμματικής και όρισε μια γλώσσα πεπερασμένης κατάστασης ως μια γλώσσα που δημιουργήθηκε από μια γραμματική πεπερασμένης κατάστασης. Αυτά τα πρώιμα μοντέλα έκαναν τις επίσημες γλώσσες ακολουθίες συμβόλων . Αυτό περιλαμβάνει τις γραμματικές χωρίς περιβάλλον, που ορίστηκαν για πρώτη φορά από τον Chomsky για φυσικές γλώσσες, αλλά ανακαλύφθηκαν ανεξάρτητα από τους Backus και Naur στην περιγραφή της γλώσσας προγραμματισμού ALGOL.

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας,
Αεροναυτικής και Περιβάλλοντος

1.3.1.2 Δεύτερη Εποχή (1957-1970)

Στη δεύτερη εποχή, η επεξεργασία γλωσσών είχε χωριστεί σε δύο παραδείγματα: συμβολική και στοχαστική. Η συμβολική δουλειά απογειώθηκε από δύο γραμμές εργασίας. Το πρώτο ήταν το έργο του Τσόμσκι, και το άλλο σχετικά με την επίσημη θεωρία της γλώσσας και τη γενετική σύνταξη στα τέλη του 50' έως τις αρχές του 60', και το έργο πολλών επιστημόνων γλωσσολογίας και υπολογιστών σχετικά με την ανάλυση αλγορίθμων, αρχικά χρησιμοποιώντας top-down και bottom-up προσεγγίσεις και μετά μέσω δυναμικού προγραμματισμού. Ένα από τα πρώτα πλήρη συστήματα ανάλυσης, που αναπτύχθηκε με βάση αυτήν την προσέγγιση ήταν το Σχέδιο Μετασχηματισμών και Ανάλυσης Ομιλίας. Η δεύτερη γραμμή έρευνας ήταν ο νέος τομέας της τεχνητής νοημοσύνης. Σε ένα εργαστήριο τεχνητής νοημοσύνης που πραγματοποιήθηκε το 1959, συζητήθηκαν τα πρώτα συστήματα φυσικής γλώσσας. Αυτά ήταν απλά συστήματα που λειτούργησαν σε έναν μόνο τομέα κυρίως μέσω ενός συνδυασμού αντιστοίχισης προτύπων και αναζήτησης λέξεων-κλειδιά με απλές ευρετικές αιτιολογίες και απαντήσεις στις ερωτήσεις. Στα τέλη της δεκαετίας του '60, είχαν αναπτυχθεί πιο επίσημα λογικά συστήματα. Στη δεκαετία του 1960 σημειώθηκε επίσης η εμφάνιση των πρώτων σοβαρών δοκιμασμένων μοντέλων επεξεργασίας ανθρώπινων γλωσσών που βασίζονται στη «γραμματική μετασχηματισμού», καθώς και στην «διαδικτυακή εταιρεία». Το Brown corpus της αμερικανικής αγγλικής ήταν μια συλλογή 1 εκατομμυρίου δειγμάτων από 500 γραπτά κείμενα από διαφορετικά είδη, (εφημερίδες, μυθιστορήματα, μη μυθοπλασία, ακαδημαϊκά κ.λπ.) που συγκεντρώθηκαν στο Πανεπιστήμιο Brown το 1963-64.

1.3.1.3 Τρίτη Εποχή (1970 – 1993)

Την επόμενη περίοδο απογειώθηκε η έρευνα στην επεξεργασία γλωσσών. Η επιτυχία του SHRDLU έδειξε ότι η ανάλυση ήταν αρκετά κατανοητή για να αρχίσει να επικεντρώνεται στη σημασιολογία και τα μοντέλα ομιλίας. Ο Roger Shank και οι συνάδελφοί του δημιούργησαν μια σειρά προγραμμάτων κατανόησης γλωσσών που επικεντρώθηκαν στην ανθρώπινη εννοιολογική γνώση, όπως σενάρια, σχέδια και στόχους, και οργάνωση ανθρώπινης μνήμης.

Τα παραδείγματα που βασίζονται στη λογική και αυτά της κατανόησης της φυσικής γλώσσας ενοποιήθηκαν σε συστήματα που χρησιμοποίησαν τη βασική λογική ως σημασιολογική αναπαράσταση. Το LUNAR είναι ένα παράδειγμα τέτοιου συστήματος.

Ένα παράδειγμα μοντελοποίησης ομιλίας έχει επικεντρωθεί σε τέσσερις βασικούς τομείς ανάλυσης ομιλίας. Η Grosz και οι συνάδελφοί της εισήγαγαν την έννοια της μελέτης της δομής στον λόγο και στην εστίαση του λόγου

1.3.1.4 Τέταρτη Εποχή (1993 -μέχρι σήμερα)

Σε αυτήν την εποχή, η χρήση πιθανολογικών και βασισμένων σε δεδομένα μοντέλων έγινε αρκετά τυπική σε όλη την επεξεργασία της φυσικής γλώσσας. Αλγόριθμοι ανάλυσης, σήμανση μέρους του λόγου, αναφορική ανάλυση και επεξεργασία λόγου, όλα άρχισαν να ενσωματώνουν πιθανότητες και να χρησιμοποιούν στρατηγικές αξιολόγησης που δανείστηκαν από την αναγνώριση ομιλίας και την ανάκτηση πληροφοριών. Επίσης, οι τεχνολογικές καινοτομίες στο hardware, όπως η αύξηση της ταχύτητας και της μνήμης των υπολογιστών, επέτρεψαν την εμπορική εκμετάλλευση ορισμένων υποτομέων επεξεργασίας λόγου και γλώσσας, ιδίως αναγνώριση ομιλίας και ορθογραφικό έλεγχο και γραμματική. Επιπλέον, η άνοδος των εφαρμογών ιστού στο πρόσφατο παρελθόν έχει προσθέσει μια νέα διάσταση της έμφασης στην ανάγκη ανάκτησης πληροφοριών βάσει γλώσσας και εξαγωγής πληροφοριών.

1.3.2 Σημαντικές εφαρμογές του NLP

- Αυτόματα σύννοψη
- Βοήθεια ανάγνωσης ξένων γλωσσών
- Βοήθεια γραφής ξένων γλωσσών
- Εξαγωγή πληροφοριών
- Ανάκτηση πληροφοριών (Information Retrieval-IR). Η IR ασχολείται με την αποθήκευση, αναζήτηση και ανάκτηση πληροφοριών. Είναι ένα ξεχωριστό πεδίο στην επιστήμη των υπολογιστών (πλησιέστερα σε βάσεις δεδομένων),

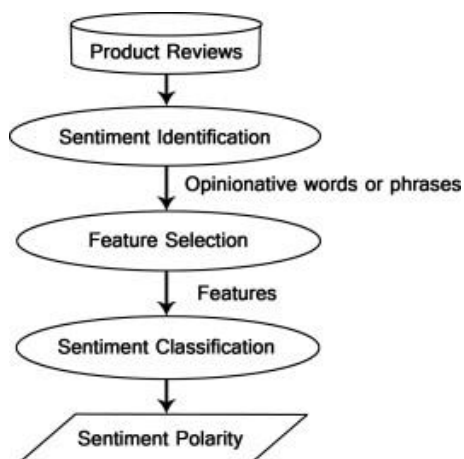
Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας, Αεροναυτικής και Περιβάλλοντος

αλλά βασίζεται σε ορισμένες μεθόδους NLP. Ορισμένες τρέχουσες έρευνες και εφαρμογές επιδιώκουν να γεφυρώσουν το χάσμα μεταξύ IR και NLP

- Machine Translating. αυτόματη μετάφραση από τη μια ανθρώπινη γλώσσα στην άλλη
- Κατανόηση φυσικής γλώσσας
- Οπτική αναγνώριση χαρακτήρων
- Απάντηση σε ερώτηση. Με δεδομένη μια ερώτηση σε ανθρώπινη γλώσσα, η διαδικασία για την παραγωγή μιας απάντησης σε ανθρώπινη γλώσσα. Η ερώτηση μπορεί να είναι κλειστή (όπως: ποια είναι η πρωτεύουσα της Ελλάδας;) ή ανοιχτή (όπως: ποια είναι η έννοια της ζωής;)
- Αναγνώριση ομιλίας. Χρησιμοποιώντας ένα ηχητικό ενός ατόμου που μιλάει, το έργο της δημιουργίας μιας υπαγόρευσης κειμένου του ομιλητή. (το αντίθετο του κειμένου σε ομιλία)
- Σύστημα διαλόγου
- Απλοποίηση κειμένου
- Κείμενο σε Ομιλία

1.4 Sentiment Analysis ή Opinion Mining

Sentiment Analysis (SA) ή Opinion Mining (OM) είναι η υπολογιστική μελέτη των απόψεων, των στάσεων και των συναισθημάτων των ανθρώπων απέναντι σε μια οντότητα. Η οντότητα μπορεί να αντιπροσωπεύει άτομα, γεγονότα ή θέματα. Αυτά τα θέματα πιθανότατα καλύπτονται από κριτικές. Οι δύο εκφράσεις (SA ή OM) εκφράζουν παραπλήσιο νόημα. Ωστόσο, ορισμένοι ερευνητές δήλωσαν ότι οι OM και SA έχουν ελαφρώς διαφορετική σημασία. Το Opinion Mining εξάγει και αναλύει τη γνώμη των ανθρώπων για μια οντότητα, ενώ το Sentiment Analysis προσδιορίζει το συναίσθημα που εκφράζεται σε ένα κείμενο και στη συνέχεια το αναλύει. Επομένως, ο στόχος της SA είναι να βρει γνώμες, να προσδιορίσει τα συναισθήματα που εκφράζουν και, στη συνέχεια, να ταξινομήσει την πολικότητά τους.

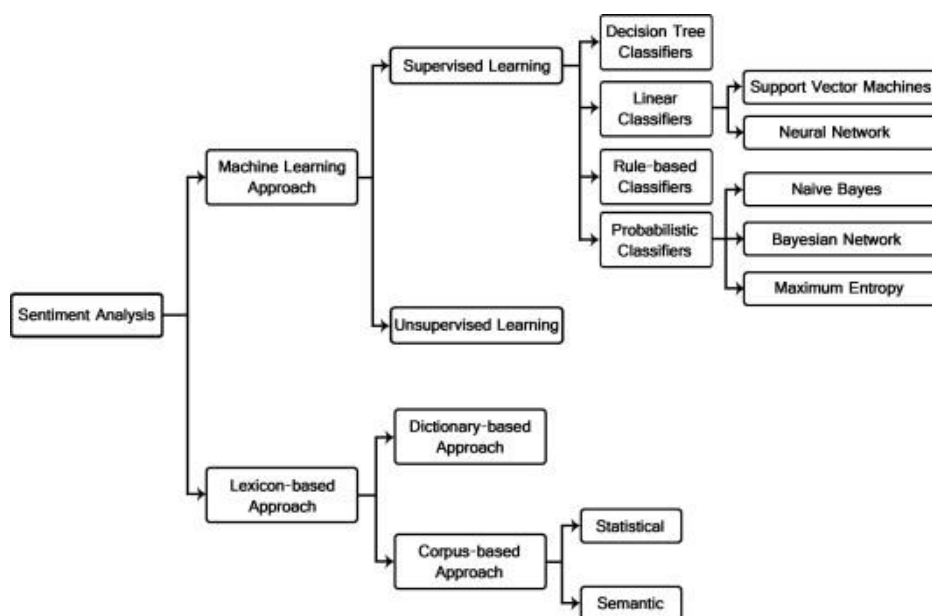


Εικόνα 1: Απεικόνιση διαδικασίας Sentiment Analysis

Η ανάλυση συναισθημάτων μπορεί να θεωρηθεί διαδικασία ταξινόμησης. Υπάρχουν τρία βασικά επίπεδα ταξινόμησης στη SA: Document-Level, Sentence-Level και Aspect-Level SA. Το Document-Level SA στοχεύει στην ταξινόμηση ενός εγγράφου γνώμης ως έκφραση θετικής ή αρνητικής γνώμης. Θεωρεί ολόκληρο το έγγραφο μια μονάδα πληροφοριών. Το Sentence-Level SA στοχεύει στην ταξινόμηση των συναισθημάτων που εκφράζονται σε κάθε πρόταση. Το πρώτο βήμα είναι να προσδιορίσουμε εάν η πρόταση είναι υποκειμενική ή αντικειμενική. Εάν η πρόταση είναι υποκειμενική, το SA θα καθορίσει εάν η πρόταση εκφράζει θετικές ή αρνητικές απόψεις. Οι Wilson et al. επεσήμαναν ότι οι εκφράσεις συναισθημάτων δεν είναι απαραίτητα υποκειμενικές. Ωστόσο, δεν υπάρχει θεμελιώδης διαφορά μεταξύ των ταξινομήσεων σε επίπεδο εγγράφου και προτάσεων, επειδή οι προτάσεις είναι απλά σύντομα έγγραφα. Η ταξινόμηση κειμένου σε Document-Level ή Sentence-Level δεν παρέχει τις απαραίτητες λεπτομέρειες των γνώμων για όλες τις πτυχές της οντότητας, οι οποίες απαιτούνται σε πολλές εφαρμογές. Για να ληφθούν αυτές οι λεπτομέρειες πρέπει να πάμε στο Aspect-Level SA. Το Aspect-level SA στοχεύει στην ταξινόμηση του συναισθήματος σε σχέση με τις συγκεκριμένες πτυχές των οντοτήτων. Το πρώτο βήμα είναι ο προσδιορισμός των οντοτήτων και των πτυχών τους. Οι κάτοχοι γνώμης μπορούν να δώσουν διαφορετικές απόψεις για διαφορετικές πτυχές της ίδιας

οντότητας, όπως αυτή η πρόταση: «Το ηγείο αυτού του τηλεφώνου δεν είναι καλό, αλλά η διάρκεια ζωής της μπαταρίας είναι μεγάλη».

Τα σύνολα δεδομένων(data sets) που χρησιμοποιούνται στο SA είναι ένα σημαντικό ζήτημα σε αυτόν τον τομέα. Οι κύριες πηγές δεδομένων προέρχονται από τις κριτικές προϊόντων. Αυτές οι κριτικές είναι σημαντικές για τους κατόχους επιχειρήσεων καθώς μπορούν να λάβουν επιχειρηματικές αποφάσεις σύμφωνα με τα αποτελέσματα ανάλυσης των απόψεων των χρηστών σχετικά με τα προϊόντα τους. Οι πηγές κριτικών είναι κυρίως ιστότοποι κριτικών. Η SA δεν εφαρμόζεται μόνο σε κριτικές προϊόντων, αλλά μπορεί επίσης να εφαρμοστεί στο χρηματιστήριο, σε άρθρα ειδήσεων, ή σε πολιτικές αντιπαραθέσεις. Στις πολιτικές αντιπαραθέσεις για παράδειγμα, θα μπορούσαμε να καταλάβουμε τις απόψεις των ανθρώπων για συγκεκριμένους εκλογικούς υποψηφίους ή πολιτικά κόμματα. Τα αποτελέσματα των εκλογών μπορούν επίσης να προβλεφθούν από πολιτικές θέσεις. Τα κοινωνικά δίκτυα και τα μικρά blog θεωρούνται πολύ καλή πηγή πληροφοριών, επειδή οι άνθρωποι μοιράζονται και συζητούν ελεύθερα τις απόψεις τους για ένα συγκεκριμένο θέμα. Χρησιμοποιούνται επίσης ως πηγές δεδομένων στη διαδικασία SA.



Εικόνα 2: Διαφορετικές Προσεγγίσεις του Sentiment Analysis

ΚΕΦΑΛΑΙΟ 2: ΣΧΕΤΙΚΗ ΕΡΕΥΝΑ (RELATED WORK)

Η παρούσα διπλωματική εργασία χρησιμοποιεί τεχνικές και μεθόδους, οι οποίες δεν είναι πρωτοποριακές και έχουν χρησιμοποιηθεί κατά την εκπόνηση και άλλων έργων. Η παρούσα εργασία έχει εμπνευστεί από αυτά τα έργα και στόχος της είναι να βελτιώσει, να επεκτείνει και κυρίως να ενώσει διαφορετικές ιδέες με σκοπό να αποφέρει καλύτερα αποτελέσματα.

Έχουν καταγραφεί πολλές εργασίες για την ανάλυση συναισθήματος (Sentiment Analysis) στη βιβλιογραφία. Μερικά από τα σημαντικότερα σχετικά έργα περιλαμβάνονται σε αυτό το κεφάλαιο. Ο Pang και Lee, χρησιμοποίησαν τεχνικές μηχανικής μάθησης για την ανάλυση συναισθήματος. Συγκεκριμένα, χρησιμοποίησαν 3 αλγόριθμους μηχανικής μάθησης: Support Maximum Entropy, Vector Machine και Naïve Bayes, οι οποίοι εφαρμόστηκαν σε ένα dataset από κριτικές ταινιών. Η ευστοχία του Naïve Bayes ήταν η υψηλότερη κατά μικρό ποσοστό. Όσον αφορά τη σχετική απόδοση, ο Naïve Bayes τείνει να χειροτερεύει και το SVM να παράγει καλύτερα αποτελέσματα κατά τη γνώμη τους, αν και οι διαφορές πάλι είναι μικρές. Πειράματα και μελέτες επιβεβαίωσαν ότι οι τεχνικές μηχανικής μάθησης ξεπέρασαν τα social-made όρια για την συναισθηματική ανάλυση σε βάσεις δεδομένων.

Για την υποστήριξη της ανάλυσης αποφάσεων μέσω της εξόρυξης γνώμης από τις κριτικές των χρηστών, έχει διερευνηθεί από πολλά έργα. Το 2004, ο Hu σχεδίασε ένα σύστημα εξόρυξης γνώμης. Χρησιμοποίησε ένα σύνολο δεδομένων του Amazon και του CNET. Χρησιμοποίησε το Wordnet για να δει αν τα εξαγόμενα επίθετα δείχνουν θετική ή αρνητική πολικότητα. Χρησιμοποίησε το Lexicon για να αναζητήσει τον γλωσσικό προσανατολισμό των επιθέτων. Ο Hu χρησιμοποίησε το πρόγραμμα ανάλυσης NLP που χρησιμοποιεί Part-Of-Speech (POS) tagging για να χωρίσει την πρόταση στα συστατικά της. Χρησιμοποίησε τους δικούς του αλγόριθμους για συχνή αναγνώριση χαρακτηριστικών, εξαγωγή λέξεων γνώμης και προσδιορισμό προσανατολισμού.

Ο Ohana δείχνει τα αποτελέσματα της χρήσης της λεκτικής πηγής SentiWordNet στο πρόβλημα της ταξινόμησης συναισθημάτων των δεδομένων που λαμβάνονται από κριτικές ταινιών. Η ανάλυσή του αξιολογεί την εφαρμογή του

SentiWordNet στην ταξινόμηση σε επίπεδο κειμένου. Η προσέγγισή του περιλαμβάνει την άθροιση θετικών και αρνητικών όρων για να αξιολογήσει τον προσανατολισμό των συναισθημάτων, και η αλλαγή εμφανίζεται με τη λήψη ενός συνόλου δεδομένων τα οποία εξασκούνται σε έναν ταξινομητή μηχανικής μάθησης. Αυτή η μελέτη δημιούργησε ευκαιρίες έτσι ώστε η γλωσσική διαδικασία να αποφέρει κέρδη στην ακρίβεια ταξινόμησης.

Άλλη μια ενδιαφέρουσα έρευνα είναι αυτή των Ajay Deshwal και Sudhir Kumar Sharma, οι οποίοι πραγματοποίησαν εξόρυξη γνώμης και Ανάλυση συναισθήματος σε κοινωνικά δίκτυα και συγκεκριμένα στο Twitter. Η συγκεκριμένη έρευνα παρουσιάζει διαφορετικά features και supervised αλγόριθμους κατηγοριοποίησης, με στόχο την σύγκριση μεταξύ τους και τον εντοπισμό της βέλτιστης μεταξύ αυτών. Παρατήρησαν ότι αρκετές ερευνητικές διαδικασίες άλλων, έδιναν βάση στην εύστοχη κατηγοριοποίηση των δημοσιεύσεων με έντονες ενδείξεις, όπως emoticons χαρούμενα ή λυπημένα, χωρίς να λαμβάνουν υπόψη τα tweets με ουδέτερο χαρακτήρα. Στην δικιά τους έρευνα, κατηγοριοποίησαν τα emoticons αλλά και τα tweets ουδέτερου χαρακτήρα. Οι αλγόριθμοι που χρησιμοποιήθηκαν είχαν τις παραμέτρους τους και δεν έκαναν fine-tuning, καθώς επίσης, σύγκριναν τις επιδόσεις του καθένα με γνώμονα την εμπειρία. Επιπλέον, απέδειξαν πως ο συνδυασμός των χαρακτηριστικών που επέλεξαν απέδωσε καλύτερα αποτελέσματα. Η βελτιστοποίηση των αλγορίθμων με την επιλογή ιδανικών παραμέτρων θα απέφερε διαφορετικά αποτελέσματα.

Η εργασία των Kanika Vishwakarma et al. είναι πολύ κοντά στην δικιά μας, όσον αφορά τον σχεδιασμό του Sentiment Analysis και τα εργαλεία που χρησιμοποιήθηκαν. Αφού συγκέντρωσαν τα Tweets, ολοκλήρωσαν την προεπεξεργασία, και στη συνέχεια πέρασαν τα δεδομένα από τον εκπαιδευμένο classifier που έφτιαξαν για τη ανάλυση. Χρησιμοποίησαν την γλώσσα προγραμματισμού Python, τις βιβλιοθήκες NLTK – Natural Language Toolkit και Scikit-learn, κατά την προεπεξεργασία των δεδομένων και στο στάδιο του Machine Learning, καθώς επίσης και το πακέτο NumPy το οποίο επιτρέπει ιδανική διαχείριση πινάκων. Σε γενικές γραμμές, οι επιδόσεις του μοντέλου τους φαίνονται αρκετά ικανοποιητικές, με μόνο μελανό σημείο το πλήθος των δεδομένων που χρησιμοποίησαν για την εκπαίδευση, 21000

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας, Αεροναυτικής και Περιβάλλοντος

κείμενα, ενώ μόλις 1400 κείμενα για δοκιμή. Το θέμα στο οποίο επικεντρώθηκαν για την συλλογή των δεδομένων τους και το dataset που χρησιμοποίησαν για να εκπαιδεύσουν τον classifier τους, δεν είναι ξεκάθαρα.

Ο Jeonghee Yi πρότεινε έναν Αναλυτή Συναισθημάτων (Sentiment Analyzer) να εξαγάγει απόψεις σχετικά με ένα θέμα από διαδικτυακά έγγραφα δεδομένων. Ο αναλυτής συναισθημάτων χρησιμοποιεί τεχνικές επεξεργασίας φυσικής γλώσσας, ανακαλύπτει όλες τις αναφορές στο θέμα και καθορίζει την πολικότητα συναισθημάτων κάθε αναφοράς. Η ανάλυση συναισθημάτων που διεξήχθη από τους ερευνητές χρησιμοποίησε τη βάση δεδομένων συναισθημάτων λεξικού (Sentiment Lexicon) και μοτίβου συναισθημάτων (Sentiment Pattern) για εξαγωγή και συσχέτιση. Το σύστημα αυτό εφαρμόστηκε σε αναφορές προϊόντων από ψηφιακές κάμερες και μουσική και τα αποτελέσματα ήταν ικανοποιητικά.

Ο Ahmed Abbasi πρότεινε νέες μεθόδους ανάλυσης συναισθημάτων για την ταξινόμηση των απόψεων σε πολλές γλώσσες. Η προτεινόμενη μέθοδος ανάλυσης συναισθημάτων χρησιμοποίησε τη λειτουργία των στιλιστικών και συντακτικών χαρακτηριστικών στην αγγλική και αραβική γλώσσα, για να αξιολογήσει τα συναισθήματα. Ο Entropy weighted Genetic Algorithm ενσωματώνεται για να βελτιώσει την απόδοση του ταξινομητή και να επιτύχει την πραγματική αξιολόγηση των κυριότερων στοιχείων. Τα πειράματα διεξήχθησαν σε κριτικές ταινιών και τα αποτελέσματα έδειξαν ότι οι προτεινόμενες τεχνικές είναι αποτελεσματικές.

ΚΕΦΑΛΑΙΟ 3: ΗΛΕΚΤΡΟΝΙΚΕΣ ΑΓΟΡΕΣ ΚΑΙ ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ

Το ηλεκτρονικό εμπόριο (E - Commerce) είναι, εκ πρώτης όψεως, ένα ηλεκτρονικό μέσο για την ανταλλαγή μεγάλων ποσοτήτων πληροφοριών προϊόντος μεταξύ χρηστών και ιστότοπων. Αυτές οι πληροφορίες πρέπει να είναι γραμμένες με σαφήνεια, καθώς οποιοσδήποτε χρήστης που έχει πρόσβαση στον ιστότοπο πρέπει να μπορεί να τις κατανοήσει. Δεδομένου του μεγάλου όγκου πληροφοριών που διατίθενται στον ιστότοπο αυτό, η αλληλεπίδραση με έναν ιστότοπο e-market γίνεται δυσκολότερη. Είναι επίσης χρονοβόρο και ο χρήστης αισθάνεται αποπροσανατολισμένος καθώς τα προϊόντα και οι πελάτες αυξάνονται συνεχώς. Μία λύση σε αυτό το πρόβλημα, η οποία θα διευκολύνει τις διαδικτυακές αγορές κατά πολύ είναι τα συστήματα σύστασης(Recommender Systems). Συστήματα προτάσεων εμφυτεύονται σε ιστότοπους ηλεκτρονικού εμπορίου για να προτείνουν υπηρεσίες και να παρέχουν στους καταναλωτές τις πληροφορίες που χρειάζονται για να αποφασίσουν για πιθανές αγορές. Αυτά τα εργαλεία λειτουργούν ως ο απόλυτα εξειδικευμένος πωλητής για κάθε πελάτη ξεχωριστά και συνήθως ενισχύονται με δυνατότητες προσαρμογής. Προσαρμόζονται στους χρήστες, βασισμένοι στην ανάλυση των προτιμήσεων και των ενδιαφερόντων τους. Τα συστήματα αυτά βασίζονται κυρίως στο προφίλ κάθε χρήστη, σε τεχνικές μάρκετινγκ και σε μεγάλες ποσότητες πληροφοριών που σχετίζονται με άλλους πελάτες και προϊόντα. Όλα αυτά γίνονται, φυσικά, σε μια προσπάθεια να προτείνουμε το σωστό αντικείμενο στον σωστό πελάτη. Εκτός αυτού, τα Recommender Systems είναι θεμελιώδη στοιχεία για τη διατήρηση της χρηστικότητας και της εμπιστοσύνης ενός ιστότοπου. Αυτός είναι και ο λόγος για τον οποίο τα site της ηλεκτρονικής αγοράς δίνουν σε αυτά σημαντικό ρόλο κατά τον σχεδιασμό τους. Εάν ένα σύστημα σύστασης θέλει να θεωρείται χρήσιμο από τους χρήστες του, πρέπει να αντιμετωπίσει διάφορα προβλήματα, όπως η έλλειψη γνώσεων που ενδεχομένως να έχουν κάποιοι χρήστες σε έναν συγκεκριμένο τομέα, η υπερφόρτωση πληροφοριών και η ελαχιστοποίηση του κόστους αλληλεπίδρασης. Τα συστήματα σύστασης του e-commerce γίνονται σταδιακά ισχυρά εργαλεία για τον χώρο των διαδικτυακών αγορών που χρησιμοποιούν πολύπλοκους μηχανισμούς κυρίως για να υποστηρίξουν τη διαδικασία λήψης αποφάσεων του χρήστη,

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας,
Αεροναυτικής και Περιβάλλοντος

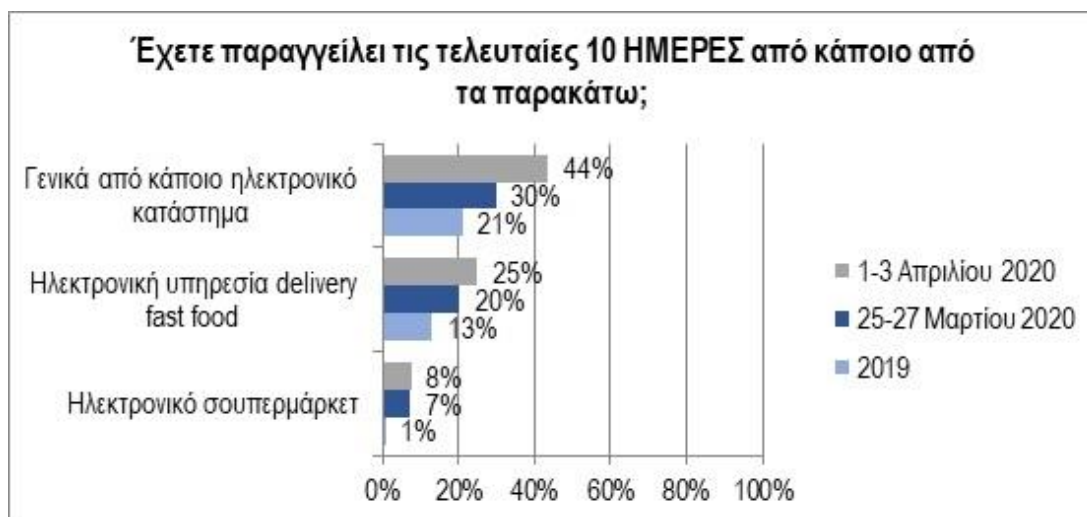
επιτρέποντας την αναλογική συλλογιστική του ανθρώπου και αποφεύγοντας τη διαδικασία αποπροσανατολισμού που συμβαίνει όταν κάποιος έχει μεγάλες ποσότητες πληροφοριών για ανάλυση και σύγκριση.

3.1 Ηλεκτρονικά καταστήματα και Ηλεκτρονικό Εμπόριο

Το ηλεκτρονικό εμπόριο άρχισε να εδραιώνεται περίπου στα μέσα του 1995 και επέφερε ραγδαίες εξελίξεις στον κόσμο των επιχειρήσεων, των αγορών καθώς και στη συμπεριφορά των καταναλωτών. Στις μέρες μας, το e-commerce μας δίνει πρόσβαση σε υπηρεσίες και δυνατότητες που στον πραγματικό κόσμο αγορών θα ήταν αδύνατες. Για παράδειγμα, καταναλωτής έχει στη διάθεσή του χιλιάδες προϊόντα, τα οποία μπορεί να συγκρίνει, ώστε να βρει το κατάλληλο για τον ίδιο. Βέβαια, αυτό είναι και καλό και κακό, διότι δημιουργείται το πρόβλημα των αμέτρητων επιλογών και οι καταναλωτές. Έτσι, το πρόβλημα αυτό καλούνται να επιλύσουν τα συστήματα συστάσεων(Recommender Systems) ώστε να διευκολύνουν ακόμα περισσότερο το χρήστη. Ως ηλεκτρονικό εμπόριο (Η.Ε.) ή ευρέως γνωστό ως e-commerce, ορίζεται το εμπόριο παροχής αγαθών και υπηρεσιών που πραγματοποιείται εξ αποστάσεως με ηλεκτρονικά μέσα, βασιζόμενο δηλαδή στην ηλεκτρονική μετάδοση δεδομένων, χωρίς να καθίσταται αναγκαία η φυσική παρουσία των συμβαλλομένων μερών, πωλητή-αγοραστή. Είναι ένα υποσύνολο της στρατηγικής «Ηλεκτρονικού επιχειρείν (e-Business)» μιας επιχείρησης και συχνά χρησιμοποιείται με σκοπό να προσθέσει ροές εισοδήματος.

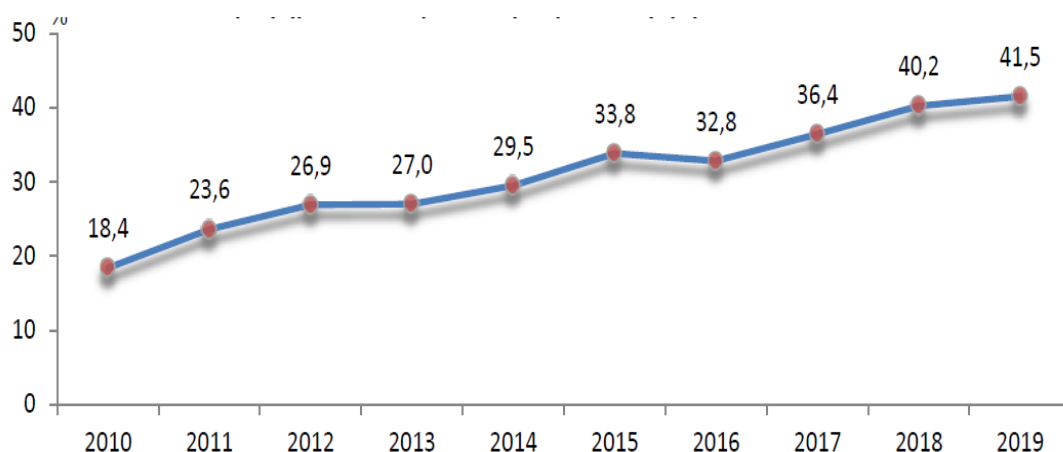
Το διαδικτυακό εμπόριο είναι ικανό να δώσει στον καταναλωτή σημαντικά περισσότερες επιλογές και ποικιλία, από ότι μπορεί το παραδοσιακό. Για παράδειγμα, ενώ ένα τυπικό κατάστημα ηλεκτρονικών παιχνιδιών(video games) προσφέρει περίπου 1500 επιλογές, το ηλεκτρονικό κατάστημα αγορών video games, Steam, προσφέρει επί του παρόντος 50000 τίτλους. Κατά συνέπεια, εξειδικευμένα προϊόντα αποκτούν σημασία στο διαδικτυακό εμπόριο ,αφού βιβλία με κατάταξη πωλήσεων κάτω των 250.000 στο Amazon.com εξακολουθούν να αντιπροσωπεύουν το 29,3% των συνολικών πωλήσεων.

Κάθε εκσυγχρονισμένη πλέον επιχείρηση έχει αναπτύξει το δικό της διαδικτυακό κατάστημα(e-shop), το οποίο κατάστημα μάλιστα αποτελεί το βασικό μηχανισμό κερδοφορίας αυτών των καταστημάτων. Σήμερα, ακόμα και η αγορά καθημερινών προϊόντων ,όπως από το super market μπορεί να γίνει ηλεκτρονικά(Amazon GO). Έρευνες του Εργαστηρίου Ηλεκτρονικού Εμπορίου (ELTURN) και του Ινστιτούτου Ερευνών Λιανεμπορίου Καταναλωτικών Αγαθών (ΙΕΛΚΑ) καταγράφουν μια σημαντική αύξηση της χρήσης του ηλεκτρονικού εμπορίου καταναλωτών. Συγκεκριμένα, τον τελευταίο χρόνο έχει διπλασιαστεί η τάση αγορών από ηλεκτρονικά καταστήματα γενικά(από το 21% του πληθυσμού που είχε κάνει τουλάχιστον μια διαδικτυακή αγορά το τελευταίο δεκαήμερο του 2019 σε 44% την πρώτη εβδομάδα του Απριλίου 2020) και πιο συγκεκριμένα στην ηλεκτρονική υπηρεσία παράδοσης έτοιμου φαγητού από 13% σε 25%. Το παραπάνω επιβεβαιώνεται και από την ανακοίνωση του Skrutz (site που φιλοξενεί καταστήματα για ηλεκτρονικές αγορές), που με 4400 ηλεκτρονικά καταστήματα που φιλοξενεί και τα 10 εκ. προϊόντα τους, κατέγραψε αύξηση των online αγορών κατά 41%. Ειδικά στις παραγγελίες από ηλεκτρονικό super market, η αύξηση των καταναλωτών είναι οκταπλάσια, αφού 8% του πληθυσμού έχει παραγγείλει online τις τελευταίες δέκα ημέρες σε σχέση με κάτω του 1% που ήταν τον 2019.



Εικόνα 3: Έρευνα ELTURN και ΙΕΛΚΑ 2020

Facebook, Instagram, YouTube, Twitter αποτελούν ιδανικά παραδείγματα της νέας μορφής ηλεκτρονικού εμπορίου για τον 21ο αιώνα. Το εναρκτήριο λάκτισμα για την ανάπτυξη του «κοινωνικού εμπορίου» έδωσε η ραγδαία εξάπλωση της κινητής τηλεφωνίας και της δικτύωσης μέσω φορητών συσκευών: μόλις το 35% των χρηστών του Facebook χρησιμοποιούν την υπηρεσία μέσω των κινητών τηλεφώνων ή των συσκευών tablet. Σύμφωνα με την ερευνά Χρήσης Τεχνολογιών Πληροφόρησης και Επικοινωνίας από Νοικοκυριά και Άτομα του έτους 2019, από την Ελληνική Στατιστική Αρχή (ΕΛΣΤΑΤ) φαίνεται ότι το 1ο τρίμηνο του 2018, το ποσοστό των χρηστών που πραγματοποίησαν ηλεκτρονικές αγορές παρουσίασε αύξηση 3,2%, ενώ υπήρξε αύξηση 125,5% την τελευταία δεκαετία(2010-2019).



Εικόνα 4: Ηλεκτρονικό εμπόριο, 1ο τρίμηνο 2010 – 2019 στην Ελλάδα.

3.2 Η Εταιρία Amazon

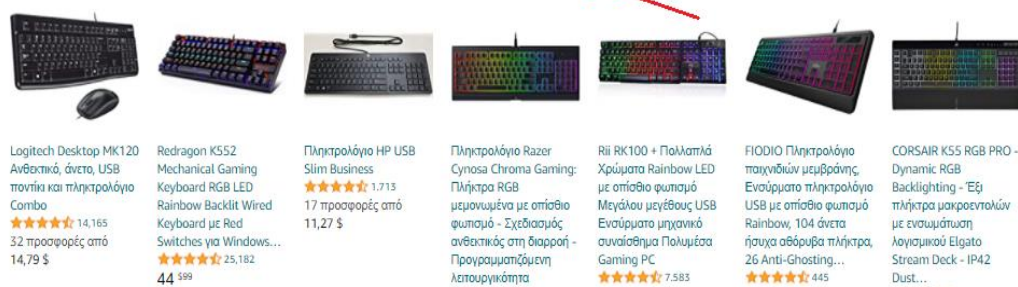
Εκτός του ότι είναι κολοσσός στον χώρο που μελετάει και κινείται η συγκεκριμένη διπλωματική εργασία, είναι και η πηγή του dataset που χρησιμοποιήσαμε και θα αναλύσουμε παρακάτω. Συνεπώς, δεν μπορούμε παρά να αναφερθούμε στην Amazon.com. Η συγκεκριμένη εταιρία, όπως είπαμε και πριν, είναι από τις μεγαλύτερες στον τομέα των ηλεκτρονικών πωλήσεων, με ένα e-shop ιδανικά σχεδιασμένο. Έδρα της είναι το Σιάτλ και ιδιοκτήτης της ο Jeff Bezos. Στις 5 Ιουλίου 1994, ο Bezos ίδρυσε αρχικά την εταιρεία στην πολιτεία της Washington με το όνομα

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας, Αεροναυτικής και Περιβάλλοντος

Cadabra. Μετά από μερικούς μήνες άλλαξε το όνομα σε Amazon.com. Στις πρώτες μέρες της, η εταιρεία λειτουργούσε έξω από το γκαράζ του σπιτιού του Bezos. Ο Μπέζος επέλεξε το όνομα κοιτάζοντας μέσα από ένα λεξικό και κατέληξε στο Amazon επειδή ήταν ένα εξωτικό και διαφορετικό μέρος, όπως οραματιζόταν για την διαδικτυακή του επιχείρηση. Η Amazon.com αρχικά λειτούργησε ως διαδικτυακό βιβλιοπωλείο και σήμερα διαθέτει τεράστια ποικιλία προϊόντων όπως ρούχα, αξεσουάρ, έπιπλα, τρόφιμα, παιχνίδια και άλλα πολλά. Η Amazon έχει αναπτύξει με τα χρόνια μια στρατηγική, η οποία επικεντρώνεται στην εξυπηρέτηση των πελατών έτσι ώστε η διαδικτυακή περιήγηση των πελατών να γίνεται ολοένα και ευκολότερη, γεγονός που αποδεικνύεται από την επένδυση της εταιρίας σε τεχνολογία και υψηλά καταρτισμένους υπάλληλους. Η εταιρεία σχεδίασε τη δική της τεχνολογία λογισμικού, όπου αποτελεί και το κύριο πλεονέκτημα της, γεγονός που επέτρεψε στην Amazon.com να εφαρμόσει μια αποτελεσματική διαχείριση σχέσεων με τους πελάτες, παρακολουθώντας τη συμπεριφορά τους κατά την περιήγηση στο site και τους επέτρεψε να κάνουν συστάσεις για παρόμοια προϊόντα. Μέσω των λειτουργιών του συστήματος αναγνώρισης χρήστη και καταγραφής της καταναλωτικής συμπεριφοράς, η εταιρεία κατάφερε να αυξήσει τις πωλήσεις κάνοντας κατάλληλες προτάσεις ή συστάσεις προϊόντων στα κατάλληλα άτομα και διαμορφώνοντας ένα φιλικό στον χρήστη περιβάλλον, το οποίο επέτρεψε την πρόσβαση και την περιήγηση σε άτομα με λίγο ή ακόμα και καθόλου γνώση των υπολογιστών.

Κατά κύριο λόγο, οι συστάσεις που πραγματοποιεί η Amazon είναι βασισμένες στο συνεργατικό φιλτράρισμα. Το συνεργατικό φιλτράρισμα(Collaborative Filtering) προτείνει στους χρήστες προϊόντα που προκύπτουν από τις αξιολογήσεις και προτιμήσεις άλλων χρηστών με παρόμοια συμπεριφορά. Για παράδειγμα, στην ιστοσελίδα της Amazon ψάχνοντας ένας πελάτης για πληκτρολόγιο, όπως φαίνεται στη παρακάτω εικόνα, γίνονται προτάσεις, σύμφωνα με τις αγορές, στη συγκεκριμένη περίπτωση, άλλων χρηστών.

Οι πελάτες που έψαξαν για "πληκτρολόγιο" τελικά αγόρασαν



Εικόνα 5: Αναζήτηση προϊόντος στην Amazon

3.3 Συστήματα Συστάσεων (Recommender Systems)

Τα Recommender Systems (RS) έχουν αναπτυχθεί εκθετικά τα τελευταία χρόνια και οι εφαρμογές αυτών έχουν εξαπλωθεί σε διάφορους τομείς της ζωής, συμπεριλαμβανομένων των ηλεκτρονικών αγορών βιβλίων, οικιακών συσκευών, ηλεκτρονικών συσκευών ταινιών, συστάσεων γιατρών και νοσοκομείων για ασθενείς, σύσταση ινστιτούτου για μαθητές και καθηγητές, προτάσεις ξενοδοχείων για τουρίστες και ούτω καθεξής. Η κεντρική ιδέα πίσω από την επιτυχία της τεχνολογίας προτάσεων είναι η τάση των ανθρώπων να βασίζονται σε εμπειρίες των γειτόνων και των φίλων τους πριν από τη λήψη οποιασδήποτε απόφασης, ειδικά όσον αφορά την αγορά οποιωνδήποτε αντικειμένων, τη συμμετοχή σε ιδρύματα τριτοβάθμιας εκπαίδευσης, επιλέγοντας διαμέρισμα προς ενοικίαση ή αγορά του, περνώντας το Σαββατοκύριακο σε κάποια σημεία διακοπών κ.λπ.

Σήμερα, το διαδίκτυο και οι δυνατότητές του έχουν σημαντικό ρόλο στην καθημερινότητά μας. Μέσα από τις μηχανές αναζήτησης ψάχνουμε να βρούμε το αντικείμενο και τις πληροφορίες που μας ενδιαφέρουν, αλλά η πληθώρα των αποτελεσμάτων που προκύπτουν δυσκολεύει την εύρεση των αντικειμένων που εξ αρχής αναζητήσαμε. Το συγκεκριμένο πρόβλημα έρχονται να λύσουν τα Συστήματα Συστάσεων (ΣΣ), τα οποία στοχεύουν στην ακριβή πρόβλεψη των προτιμήσεων και των ενδιαφερόντων των χρηστών μέσω της δημιουργίας προτάσεων. Ως συστήματα συστάσεων ορίζονται τα συστήματα όπου οι άνθρωποι παρέχουν προτάσεις ως εισόδους, οι οποίες στη συνέχεια συγκεντρώνονται και κατευθύνονται σε συγκεκριμένους παραλήπτες. Τα αναγκαία δεδομένα συλλέγονται με δύο τρόπους:

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας, Αεροναυτικής και Περιβάλλοντος

1. είτε μέσα από αξιολογήσεις(ratings) των χρηστών
2. είτε κάνοντας tracking το ιστορικό κάθε επισκέπτη

Στη συνέχεια, τα δεδομένα αυτά συγκρίνονται με εκείνα που έχουν προκύψει από άλλους χρήστες, επεξεργάζονται και σύμφωνα με τις ομοιότητες που παρουσιάζουν προκύπτουν οι κατάλληλες προτάσεις, στους κατάλληλους χρήστες.

Η πρόοδος των τεχνολογιών του Διαδικτύου προκάλεσε υπερφόρτωση δεδομένων, με αποτέλεσμα οι αγοραστές να αντιμετωπίζουν περισσότερες δυσκολίες στην εύρεση του ακριβούς προορισμού που ικανοποιεί τις ανάγκες τους μιας και η συλλογή από τις διαθέσιμες επιλογές είναι τεράστια. Για παράδειγμα, αν ένας μαθητής επιθυμεί να περάσει τις διακοπές του σε ένα ορεινό μέρος με γαλήνη και ηρεμία, θα του εμφανίζονταν χιλιάδες μέρη σε όλο τον κόσμο (κατά την αναζήτηση) με αυτά τα χαρακτηριστικά. Σε μια τέτοια περίπτωση, τα Συστήματα Συστάσεων μπορούν να παρέχουν μια καλύτερη επιλογή ανάλογα με την ανάγκη και την απαίτηση του κάθε χρήστη και ανάλογα με τις προηγούμενες προτιμήσεις του.

Τα Συστήματα Συστάσεων πλέον είναι βασικό και κομβικό κομμάτι στην στρατηγική και των εταιρειών, αφού μέσω των εξατομικευμένων προτάσεων προς τους καταναλωτές πετυχαίνουν την ικανοποίηση των πελατών τους, πράγμα που σημαίνει την αύξησή τους, άρα και την αύξηση των εσόδων της εκάστοτε εταιρείας. Ανάμεσα στις πιο δημοφιλείς ιστοσελίδες που χρησιμοποιούν συστήματα συστάσεων είναι η Amazon.com, η οποία παρέχει μια εξατομικευμένη ιστοσελίδα για κάθε μεμονωμένο χρήστη, το Netflix που χρησιμοποιεί εξατομικευμένα συστήματα για να προταθούν ταινίες και τηλεοπτικές εκπομπές, όπως και το Facebook το οποίο προτείνει πιθανούς φίλους στους χρήστες. Ένας άλλος παράγοντας που οι επιχειρήσεις χρησιμοποιούν τα συστήματα συστάσεων είναι η αύξηση της αξιοπιστίας. Για τη δημιουργία εξατομικευμένων συστάσεων χρειάζεται η δημιουργία προφίλ χρηστών καθώς αυτά αποτελούν αντιπροσωπευτικές αναπαραστάσεις των χαρακτηριστικών και αναγκών των χρηστών.

Παρόλο που υπάρχουν διάφοροι ορισμοί που έχουν προτείνει οι ερευνητές για τα συστήματα σύστασης, εμείς τα ορίζουμε ως: "Τα Συστήματα Συστάσεων

προσπαθούν να προσδιορίσουν την ανάγκη και τις προτιμήσεις των χρηστών, να φιλτράρουν αναλόγως την τεράστια συλλογή δεδομένων και να παρουσιάσουν

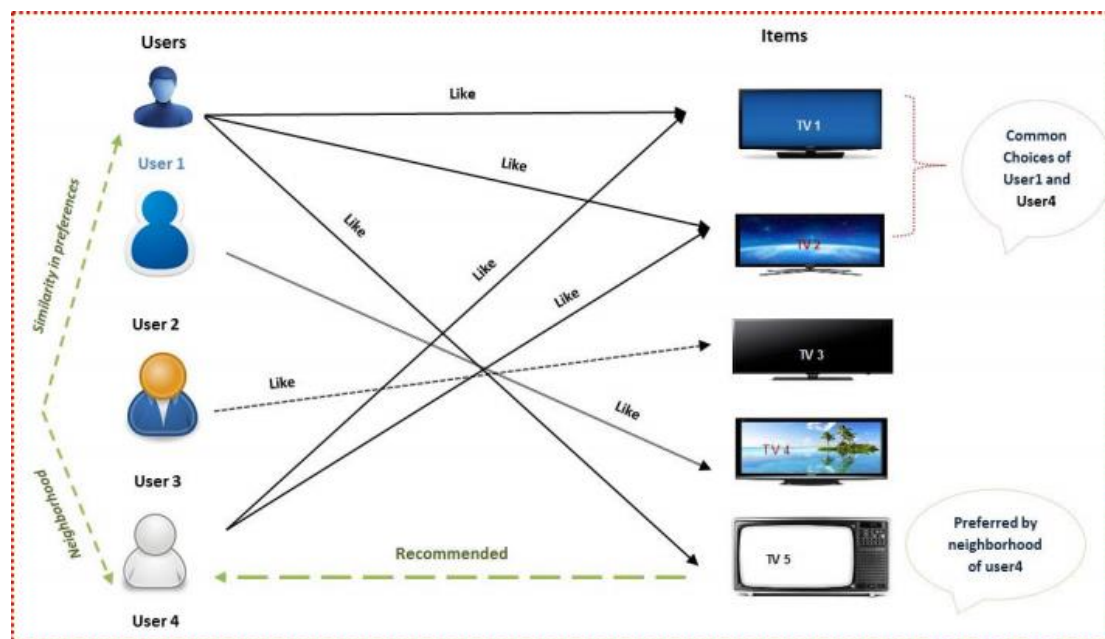
3.4 Κατηγορίες Συστημάτων Σύστασης

Τα συστήματα συστάσεων μπορούν να κατηγοριοποιηθούν σε διάφορες βάσεις. Σε αυτήν την διπλωματική, έχουμε κατηγοριοποιήσει 8 τύπους συστημάτων σύστασης σύμφωνα με το {10} και θα αναλύσουμε τις 5 συνηθέστερες. Αυτές οι κατηγορίες καλύπτουν γενικά τις τεχνικές που έχουν χρησιμοποιηθεί από τις μάζες ή έχουν εφαρμόσει ερευνητές της τρέχουσας γενιάς.

3.4.1 Collaborative Filtering Recommender Systems

Είναι η πιο επιτυχημένη και συχνά χρησιμοποιούμενη τεχνική συστάσεων που συζητήθηκε στη βιβλιογραφία από την εμφάνιση του πρώτου συστήματος συστάσεων στα μέσα της δεκαετίας του 1990. Η συνεργατική προσέγγιση χρησιμοποιεί τη σύσταση από άλλους πελάτες των οποίων οι επιλογές είναι παρόμοιες με τους πελάτες-στόχους. Οι πελάτες με παρόμοιες επιλογές ονομάζονται γείτονες.

Έτσι, δύο σημαντικές εργασίες εκτελούνται στο συλλογικό φιλτράρισμα: 1) εύρεση των γειτόνων ενός στόχου και 2) διερεύνηση των προτιμήσεων των γειτόνων ενός πελάτη ή χρήστη. Ο γείτονας ενός χρήστη μπορεί να σχηματιστεί αναλύοντας την προηγούμενη αγοραστική συμπεριφορά του χρήστη και υπολογίζοντας τις βαθμολογίες ομοιότητας μεταξύ των επιλογών αυτών των χρηστών. Η σύσταση των πελατών μιας «γειτονιάς» μπορεί να ληφθεί είτε ρητά (ως προς τη βαθμολογία που είναι αριθμητικές τιμές εντός ενός καθορισμένου εύρους), είτε έμμεσα με ορισμένα καθορισμένα μέτρα. Οι έμμεσες συστάσεις περιλαμβάνουν επίσης τα σχόλια του πελάτη. Τα σχόλια του πελάτη είναι η συμπεριφορά που παρατηρείται από τα στοιχεία καταγραφής του χρήστη ή μπορεί να είναι τα συναισθήματα των χρηστών που εκφράζονται κατά τις κριτικές τους.



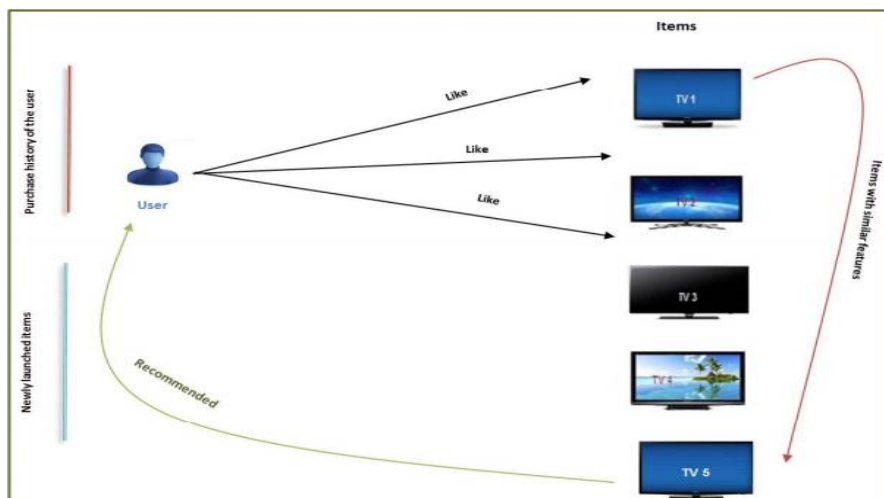
Εικόνα 6: Collaborative Filtering Recommender Systems

3.4.2 Content-Based Recommender Systems

Είναι σαφές από την παραπάνω συζήτηση ότι το συνεργατικό φιλτράρισμα βασίζεται στην εύρεση ομοιοτήτων μεταξύ των χρηστών. Σε αντίθεση με το συλλογικό φιλτράρισμα, η content-based προσέγγιση εκμεταλλεύεται τα χαρακτηριστικά του εκάστοτε προϊόντος και τα χρησιμοποιεί για να κάνει προτάσεις. Οι μέθοδοι πρότασης με βάση το περιεχόμενο θεωρούνται συμπληρωματικές των συνεργατικών τεχνικών και δίνουν έμφαση στην εύρεση ομοιότητας μεταξύ αντικειμένων.

Ας εξετάσουμε το παράδειγμα που απεικονίζεται χρησιμοποιώντας το Σχ. 2. Υπάρχουν πέντε διαφορετικές τηλεοράσεις για τις οποίες περιγράφεται μια προσέγγιση βάσει περιεχομένου για έναν χρήστη. Ο χρήστης έχει προτιμήσει τις TV1, TV2 και TV3 αγοράζοντάς τις. Οι TV4 και TV5 είναι νέα είδη. Τα χαρακτηριστικά της TV5

είναι παρόμοια με της TV1, ενώ η TV4 έχει διαφορετικά χαρακτηριστικά. Έτσι, η προσέγγιση βάσει περιεχομένου θα συνιστούσε την TV5 στον χρήστη και όχι την TV4.



Εικόνα 7 : Content-Based Recommender System

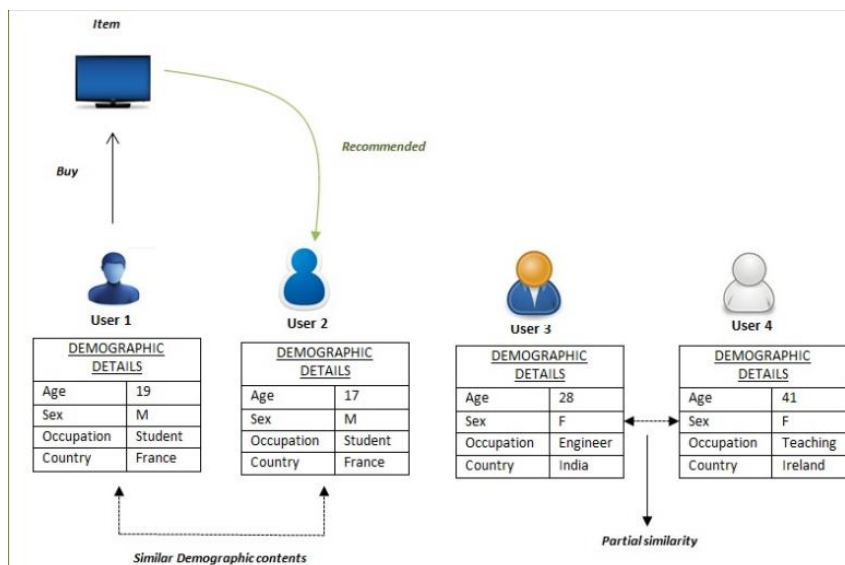
Η σύσταση βάσει περιεχομένου προήλθε κυρίως από την έννοια της πληροφόρησης και είναι ένα είδος μεθόδου πρότασης που βασίζεται στη σύγκριση των προτιμήσεων των χρηστών και στη συσχέτιση του περιεχομένου μεταξύ των ονομάτων με σκοπό την παροχή συστάσεων στους χρήστες. Αυτή η μέθοδος κρίνει και ανακαλύπτει αντικείμενα που πιθανώς ενδιαφέρουν τους χρήστες αναλύοντας τα χαρακτηριστικά βάσει του Προφίλ κάθε χρήστη. Τα αποτελέσματα στη συνέχεια συνιστώνται στους χρήστες.

3.4.3 Demographic Filtering based Recommender Systems

Τα συστήματα σύστασης που βασίζονται στο δημογραφικό φιλτράρισμα χρησιμοποιούν επίσης κριτήρια ομοιότητας ως μετρικό σύστημα. Ωστόσο, αντί να βρει παρόμοια στοιχεία από γειτονικούς χρήστες, προσπαθεί να βρει την ομοιότητα μεταξύ των δημογραφικών πληροφοριών των χρηστών, όπως, ηλικία, φύλο, επάγγελμα κ.λπ. Σε αυτήν την προσέγγιση, το σύστημα αποθηκεύει τις δημογραφικές πληροφορίες των πελατών και όποτε έρχεται ένας νέος χρήστης στον ιστότοπο για την αγορά οποιουδήποτε προϊόντος, το σύστημα προσδιορίζει την ομοιότητα μεταξύ των δημογραφικών πληροφοριών των χρηστών. Σύμφωνα με την προτίμηση του πελάτη,

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας, Αεροναυτικής και Περιβάλλοντος

το σύστημα προτείνει όμοια αντικείμενα σε έναν νέο χρήστη που έχει παρόμοια ηλικία, φύλο, επάγγελμα κλπ με έναν άλλον πελάτη. Μια τυπική προσέγγιση προτάσεων του συστήματος συστάσεων βάσει δημογραφικού φιλτραρίσματος παρουσιάζεται πιο κάτω



Εικόνα 8: Demographic based Recommender System

3.4.4 Knowledge based Recommender Systems

Τα συγκεκριμένα συστήματα λειτουργούν πάνω στη γνώση που έχουν για κάποια συγκεκριμένα χαρακτηριστικά είτε του αντικειμένου είτε του χρήστη και με γνώμονα αυτά κάνουν τις κατάλληλες προτάσεις στους κατάλληλους καταναλωτές. Για τη λειτουργία αυτών των συστημάτων χρειάζονται γνώσεις που σχετίζονται άμεσα με τον χρήστη, άμεσα με τα αντικείμενα και γύρω από το πως οι ανάγκες κάθε χρήστη μπορούν να καλυφθούν από τα συγκεκριμένα αντικείμενα. Τα συστήματα που βασίζονται στην γνώση προσαρμόζονται άμεσα στις αλλαγές των προτιμήσεων των χρηστών, χάρις την ικανότητά τους να αντλούν πληροφορίες για τους χρήστες συνεχώς. Στην πραγματικότητα τα συστήματα που βασίζονται στη γνώση παρουσιάζουν αρκετές ομοιότητες με αυτά που βασίζονται στο περιεχόμενο. Σημαντική διαφορά των δύο είναι ότι τα content-based συστήματα μαθαίνουν από τις προηγούμενες δράσεις του χρήστη, ενώ τα knowledge-based συστήματα λειτουργούν με τις προδιαγραφές του χρήστη, των

αναγκών και των συμφερόντων τους. Τα συστήματα βασισμένα στη γνώση χωρίζονται σε δυο βασικές κατηγορίες: 1) Constraint-based recommender systems τα οποία βασίζονται σε κάποιους περιορισμούς που προκύπτουν από τις απαιτήσεις και τα χαρακτηριστικά του χρήστη και 2) Case-based recommender systems όπου χρησιμοποιούν μέτρα ομοιότητας για να μπορέσουν να κάνουν τις συστάσεις που επιθυμούν.

Σε γενικές γραμμές, τα συστήματα συστάσεων που λειτουργούν με βάση τη γνώση, είναι κατάλληλα για λιγότερο δημοφιλή προϊόντα, που δεν αγοράζονται συχνά. Βασικό τους πλεονέκτημα είναι ότι έχουν την δυνατότητα να κάνουν εύστοχες και ακριβείς προτάσεις σε χρήστες, των οποίων οι επιλογές δεν συνάδουν με το μεγαλύτερο κομμάτι του αγοραστικού κοινού.

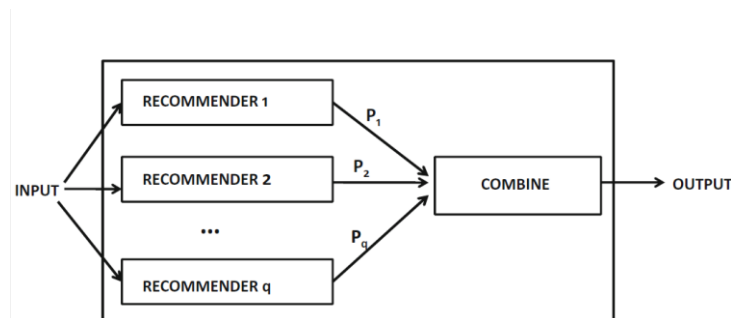
3.4.5 Hybrid Recommender Systems

Παρόλο που το Συνεργατικό φιλτράρισμα και αυτό του περιεχομένου αποτελούν τις πιο συχνά χρησιμοποιούμενες τεχνικές στο σχεδιασμό Συστημάτων Πρότασης, δεν παρέχουν επαρκείς εξηγήσεις σχετικά με την αιτία των συγκεκριμένων προτάσεων σε συγκεκριμένους χρήστες, επομένως, αποτυγχάνουν να εξηγήσουν τα αίτια κάποιων προτάσεων σε διάφορα σενάρια. Αυτές οι αδυναμίες και των δύο κορυφαίων τεχνολογιών μπορούν να ξεπεραστούν με τη χρήση του συνδυασμού αυτών. Αυτοί οι συνδυασμοί ονομάζονται «υβριδική τεχνική». Έχουμε κατηγοριοποιήσει επτά τύπους υβριδικών συστημάτων σύστασης βάσει διαφορετικών συνδυασμών, σύμφωνα με το [10].

- Σταθμισμένα συστήματα συστάσεων: Τα αποτελέσματα των συνιστώμενων συστάσεων αθροίζονται με γραμμικό συνδιασμό για τον υπολογισμό βαθμολογιών.
- Συστήματα συστάσεων εναλλαγής: Ο αλγόριθμος εναλλάσσεται μεταξύ διαφόρων συστημάτων σύστασης ανάλογα με τρέχουσες ανάγκες.
- Συστήματα Συστάσεων Cascade: Το συγκεκριμένο σύστημα σε πρώτη φάση παράγει μια ευρύτερη κατάταξη των ενδιαφερόμενων και σε δεύτερη φάση βελτιώνει τη σύστασή τους.

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας, Αεροναυτικής και Περιβάλλοντος

- **Αύξηση Χαρακτηριστικών:** Η έξοδος ενός συστήματος προτάσεων αξιοποιείται για να δημιουργήσει μια είσοδο χαρακτηριστικών για το επόμενο. Με αυτό τον τρόπο βελτιώνεται το βασικό σύστημα.
- **Συνδυασμός χαρακτηριστικών:** Τα χαρακτηριστικά συνδυάζονται από διαφορετικές πηγές δεδομένων και χρησιμοποιούνται στα πλαίσια ενός συστήματος σύστασης.
- **Μικτά υβριδικά συστήματα συστάσεων:** Σε περίπτωση που είναι αναγκαίο να γίνει μεγάλος όγκος προτάσεων ταυτόχρονα, με τη χρήση αυτής της τεχνικής μπορούν να παρουσιαστούν μαζί όλες οι συστάσεις από μια ή περισσότερες τεχνικές.
- **Meta-level:** Το μοντέλο που χρησιμοποιείται από ένα σύστημα σύστασης, χρησιμοποιείται ως είσοδος σε άλλο.



Εικόνα 9: Hybrid Recommender System

ΚΕΦΑΛΑΙΟ 4: ΒΙΒΛΙΟΓΡΑΦΙΑ

- {1} Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, “Thumbs up? Sentiment classification using machine learning techniques”, In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86, 2002.
- {2} B. Pang and L. Lee, “A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts,” 2004.
- {3} Ahmed Abbasi, Hsinchun Chen, And Arab Salem, “Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums”, ACM Trans. Inf. Syst., Vol. 26, No. 3. (June 2008), pp. 1-34.
- {4} Kanika Vishwakarma Goldi Rawat Priyanka Badhani Bhumika Gupta, Monika Negi. Study of twitter sentiment analysis using machine learning algorithms on python. International Journal of Computer Applications, 2017
- {5} Sudhir Kumar Sharma Ajay Deshwal. Twitter sentiment analysis using various classification algorithms. Conference: 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2016.
- {6} B. Ohana and B. Tierney, “Sentiment classification of reviews using SentiWordNet,” Sch. Comput. 9th. IT T Conf., p. 13, 2009.
- {7} M. Hu and B. Liu, “Mining and summarizing customer reviews,” Proc. 2004 ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '04, p. 168, 2004.
- {8} M. Hu and B. Liu, “Mining Opinion Features in Customer Reviews,” 19th Natl. Conf. Artificial Intell., pp. 755–760, 2004.
- {9} Yi, J., T. Nasukawa, R. Bunescu, and W. Niblack: 2003, “Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques”, In: Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM-2003). Melbourne, Florida.
- {10} Shahab Saquib Sohail, Jamshed Siddiqui and Rashid Ali, Review Article, “Classifications of Recommender Systems: A review
- {11} Jason Brownlee in Machine Learning Algorithms, ‘Logistic Regression for Machine Learning’

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας,
Αεροναυτικής και Περιβάλλοντος

- {11} Αργυροπούλου Χριστοδούλου, Διπλωματική εργασία, “Σχεδιασμός και υλοποίηση εφαρμογής για τη συγκέντρωση και ανάλυση δεδομένων που αφορούν την κυβερνοασφάλεια μέσω του γνωστού μέσου κοινωνικής δικτύωσης Twiter”
- {12} Nisha Jebaseeli, E. Kirubakaran, “A Survey on Sentiment Analysis of (Product) Reviews
- {13} Momina Shaheen et al., “Sentiment Analysis on Mobile Phone Reviews Using Supervised Learning Techniques”
- {14} M. IKONOMAKIS, S. KOTSIANTIS, V.TAMPAKAS, “TEXT CLASSIFICATION USING MACHINE LEARNING TECHNIQUES”
- {15} Μιχαλοπούλου Μαρία, Διπλωματική Εργασία, “Δημιουργία Υβριδικού Συστήματος Παροχής Συστάσεων με την Αξιοποίηση Τεχνικών Ανάλυσης Συναισθήματος (Sentiment Analysis) των Αξιολογήσεων Χρηστών σε Ηλεκτρονικές Συναλλαγές”.
- {16} Aggarwal., Charu C. *Recommender Systems*. s.l. : Springer, 2016.
- {17} Dionisis Margaritis, Dionysios Vasilopoulos, Costas Vassilakis and Dimitris Spiliotopoulos. *Improving Collaborative Filtering’s Rating Prediction Coverage in Sparse Datasets through the Introduction of Virtual Near Neighbors*.
- {18} Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*.
- {19} Sandra Vieira¹, Walter Hugo Lopez Pinaya, Andrea Mechelli. *Introduction to machine learning*
- {20} Debbie Zhang, Simeon Simoff*, Silvana Aciar, John Debenham. *A multi agent recommender system that utilises consumer reviews in its recommendations*. 2008.
- {21} Aggarwal, Charu C. *Data Mining: The Textbook*. 2015.
- {22} BURKE, ROBIN. *Hybrid Recommender Systems: Survey and Experiments*. 2002.
- {23} Nikos Kanakaris, Nikos Karacapilidis, Georgios Kournetas and Alexis Lazanas(. *Combining Machine Learning and Operations Research Methods to Advance the Project Management Practice*. University of Patras. 2020.
- {24} Eleanor Mulholland, Paul Mc Kevitt, Tom Lunney, and Karl-Michael Schneider. *Analysing Emotional Sentiment in People’s YouTube Channel Comments*. 2017.

Τμήμα Μηχανολόγων και Αεροναυπηγών Μηχανικών – Τομέας Ενέργειας, Αεροναυτικής και Περιβάλλοντος

- {25} Li Chen Cheng, Ming-Chan Lin. *A hybrid recommender system for the mining of consumer preferences from their reviews*. 2019.
- {26} Tushar Ghorpade, Lata Ragha. *Featured Based Sentiment Classification for Hotel Reviews using NLP and Bayesian Classification*. India : s.n., 2012.
- {27} Richa Sharma, Shweta Nigam and Rekha Jain. *Supervised Opinion Mining Techniques: A Survey*. 2013
- {28} Akkamahadevi R Hanni, Mayur M Pati. *Summarization of Customer Reviews for a Product on a website using Natural Language Processing*. 2017.
- {29} Mich, L. *From Natural Language to Object Oriented Using the Natural Language Processing System LOLITA*. 1996.
- {30} M. Rajman, R. Besancon. *Text Mining:Natural Language techniques and Text Mining applications*,1998.
- {31} Tushar Ghorpade, Lata Ragha. *Featured Based Sentiment Classification for Hotel Reviews using NLP and Bayesian Classification*. 2012.
- {32} Hsaio-Fan Wang, Cheng-Ting Wu. *A Strategy-Oriented Operation Module for Recommender Systems in E-Commerce*. Department of Industrial Engineering and Engineering Management. CHINA : s.n.
- {33} R.D. Lawrence, G.S. Almasi, V. Kotlyar, M.S. Viveros, S.S. Duri. *Personalization of supermarket product recommendations*,. 2001.
- {34} Oliver Hinz & Jochen Eckert. *The Impact of Search and Recommendation Systems on Sales in Electronic Commerce*. 2010
- {35} Παπάζογλου, Νίκη. *insider.gr*. [Ηλεκτρονικό] 18 Αυγούστου 2019 <https://www.insider.gr/specials/foyntonei-i-mahi-toy-online-delivery>.
- {36} APXH, ΕΛΛΗΝΙΚΗ ΣΤΑΤΙΣΤΙΚΗ. *statistics.gr*. [Ηλεκτρονικό] 11 Δεκεμβρίου 2019. <https://www.statistics.gr/documents/20181/1fb15bd8-a8b8-cbe4-dc66-e8574a3df1c0>.
- {37} Filson, Darren. *The Impact of E-Commerce Strategies on Firm Value : Lessons from Amazon.com and its Early Competitors*. 2003.
- {38} Aggarwal., Charu C. *Recommender Systems*. s.l. : Springer, 2016.

{39} Χρήστος Π. Κωτσαλένης, Μεταπτυχιακή Διπλωματική Εργασία, Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών «Μαθηματικά των Υπολογιστών και των Αποφάσεων», “Χρήση Τεχνικών Μηχανικής Μάθησης για την Επιλογή Βέλτιστου Αλγόριθμου για την Επίλυση Αραιών Γραμμικών Συστημάτων”

{40} Alexander L. Frdkov, Institute for Problems in Mechanical Engineering, “Early History of Machine Learning”