

Πώς φτάσαμε στο *ChatGPT* και τι μπορούμε να κάνουμε μαζί του

Στέργιος Χατζηκυριακίδης
Πανεπιστήμιο Κρήτης

7 Μαρτίου 2024

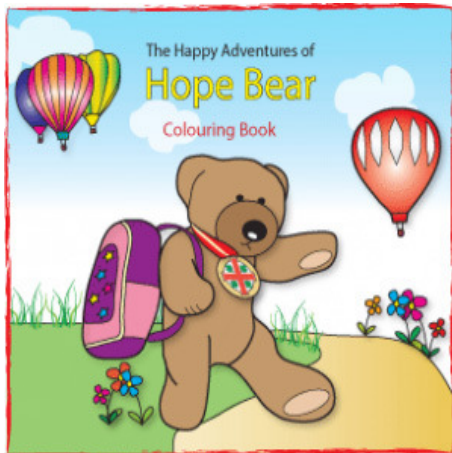
- Ή αλλιώς, ο φίλος μου ο υπολογιστικός μοντέλος!
 - Στην αρχή που τα ελληνικά του μοντέλου δεν ήταν ακόμα ιδιαίτερα καλά, έτσι μου είχε συστηθεί, ως υπολογιστικός μοντέλος! Το κράτησα από τότε

- Ας πάμε λίγο πίσω, το ChatGPT δεν φύτρωσε ξαφνικά

- Αποτέλεσμα σωρευτικής έρευνας σε ένα αντικείμενο που ονομάζεται Υπολογιστική Γλωσσολογία, ή Επεξεργασία Φυσικής Γλώσσας
 - Το αντικείμενο μετρά ήδη 7 δεκαετίες έρευνας
- Χοντρικά: η τομή των επιστημών της Γλωσσολογίας και της Επιστήμης της Πληροφορίας (ανάλογα την περίπτωση, αλλά και την ερευνητική εποχή, τείνει να γέρνει προς μία από τις επιμέρους κατευθύνσεις περισσότερο)

- Πέρασμα από συστήματα κανόνων, οι οποίοι δίνονται από ειδικούς σε συστήματα που μαθαίνουν τους κανόνες από τα δεδομένα
 - Συμβολικά συστήματα vs συστήματα Μηχανικής Μάθησης

- Πώς ξεκίνησαν όλα...

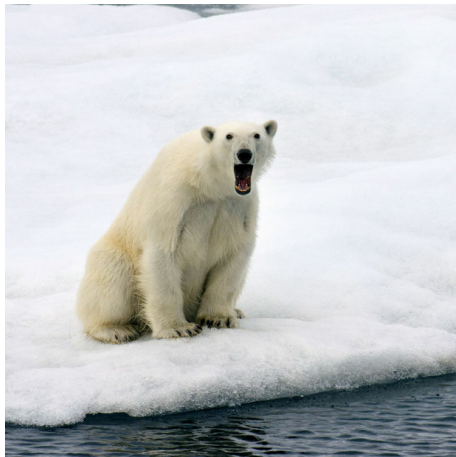


- Συνενώνοντας έναν συμβολικό συντακτικό αναλυτή με λογική σημασιολογία
 - Καθορισμός αντιστοιχίας μεταξύ αφηρημένων συντακτικών δομών (συνήθως αφηρημένων συντακτικών δέντρων) και σημασιολογίας εκπεφρασμένη σε κάποιο λογικό σύστημα

- Εξαιρετικά πετυχημένη σε ελεγχόμενα περιβάλλοντα
 - Πολύ ακριβής και λεπτομερής (fine-grained)



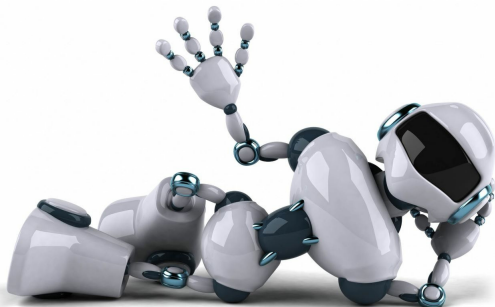
- Δεν τα πάει καλά σε ανοιχτό κείμενο ή γενικότερα μη ελεγχόμενα περιβάλλοντα



- Αντικατάσταση από μεθόδους Μηχανικής Μάθησης



- Τέλη δεκαετίας του '90, μέχρι το 2013-2014 περίπου
 - Εποπτευόμενη μηχανική μάθηση: χρήση handmade χαρακτηριστικών για την εκπαίδευση των μοντέλων
 - Bag-of-Words, Similarity Metric Based Models , Maximum Entropy Classifiers, SVMs, Naive Bayes... οι συνήθεις ύποπτοι
 - Αρκετά με τη Λογική! Αυτό θα δουλέψει σίγουρα!



Βήμα 1: Προετοιμασία Δεδομένων

Διαχωρίστε το σύνολο δεδομένων σε δύο μέρη:
εκπαίδευση και δοκιμή.

ˆ Βήμα 2: Προεπεξεργασία Δεδομένων

Για κάθε μέιλ στο σύνολο δεδομένων:

Μετατροπή κειμένου σε πεζά.

Χωρισμός κειμένου σε λέξεις.

Προεπεξεργασία

Καταμέτρηση συχνότητας κάθε λέξης.

ˆ Βήμα 3: Υπολογισμός Πιθανοτήτων

Υπολογισμός πρότερων πιθανοτήτων για σπαμ και μη σπαμ.

Για κάθε λέξη:

Υπολογισμός πιθανότητας δεδομένου του σπαμ.

Υπολογισμός πιθανότητας δεδομένου του μη σπαμ.

^ Βήμα 4: Λειτουργία Ανίχνευσης Σπαμ

Ορισμός συνάρτησης για ταξινόμηση νέων μέιλ σε σπαμ ή μη σπαμ.

^ Βήμα 5: Δοκιμή του Ταξινομητή

Για κάθε μέιλ στα δεδομένα δοκιμής:

Χρήση της συνάρτησης ανίχνευσης σπαμ για ταξινόμηση.

^ Βήμα 6: Αξιολόγηση Επίδοσης

Υπολογισμός μετρικών επίδοσης.

«Ένα τεχνητό νευρωνικό δίκτυο (ΤΝΔ) είναι ένα σύστημα επεξεργασίας πληροφοριών που εμπνέεται από τον τρόπο που τα βιολογικά νευρικά συστήματα, όπως ο εγκέφαλος, επεξεργάζονται πληροφορίες. Το βασικό στοιχείο αυτού του συστήματος είναι η καινοτόμα δομή του συστήματος επεξεργασίας πληροφοριών. Αποτελείται από έναν μεγάλο αριθμό διασυνδεδεμένων στοιχείων επεξεργασίας (νευρώνων) που εργάζονται από κοινού για την επίλυση συγκεκριμένων προβλημάτων. Τα ΤΝΔ, όπως και οι άνθρωποι, μαθαίνουν μέσω παραδειγμάτων. Ένα ΤΝΔ «κουρδίζεται» για μια συγκεκριμένη εφαρμογή μέσω μιας διαδικασίας μάθησης. Η μάθηση στα βιολογικά συστήματα περιλαμβάνει προσαρμογές στις συναπτικές συνδέσεις που υπάρχουν μεταξύ των νευρώνων. Αυτό ισχύει και για τα ΤΝΔ.»

Αυτόματη μετάφραση από τα Αγγλικά με ελάχιστες τροποποιήσεις, από τον μεταφραστή της Γκουγκλ, ο οποίος είναι πλέον νευρομεταφραστής!

- Κυρίαρχη προσέγγιση με διάφορα μοντέλα βαθείας μάθησης
 - RNNs, LSTMs
 - pretrained models (large language models, BERTology)

Μεγάλα γλωσσικά μοντέλα

- Η περίπτωση του περίφημου BERT
- Μάλλον είναι το πιο πολυχρησιμοποιημένο μοντέλο στην ΕΦΓ τα τελευταία 3 με τέσσερα χρόνια (πλέον κονταροχτυπιέται με το GPT)



- Ανήκει στη γενικότερη ομάδα μοντέλων νευρωνικών δικτύων, που λέγονται μετασχηματιστές

- Τι κάνει το μοντέλο BERT;
 - Μεταφορά μάθησης: μεταφορά γνώσεων που λαμβάνονται από τη μία εργασία στην άλλη
 - Αυτο-προσοχή (self-attention): απλοϊκά, ένας μηχανισμός που επιτρέπει το μοντέλο να κοιτάει τον περίγυρο του για να καταλάβει καλύτερα την εκάστοτε λέξη. Όλες οι λέξεις συσχετίζονται μεταξύ τους και στο τέλος δίνεται ένα σκορ της πρότασης
 - Προεκπαίδευση του μοντέλου σε δύο εργασίες πρόβλεψης (η μία είναι π.χ. προβλέψτε την επόμενη λέξη)
 - Βελτιστοποίηση για άλλες εργασίες: δεν απαιτείται εκπαίδευση από το μηδέν, απλώς προσαρμογή/κούρδισμα (fine-tuning)
 - Εντυπωσιακά αποτελέσματα σε σωρεία εργασιών Υπολογιστικής Γλωσσολογίας

- Βασίζεται στο πρότερο μεγάλο γλωσσικό μοντέλο με την ονομασία GPT-3.5 (τώρα είμαστε πλέον στο 4, και στο εργαστήριο στο 5)

- Είναι ευφυές;
 - Δεν είμαι πολύ σίγουρος ότι ξέρω τι ακριβώς είναι η ευφυΐα, για να σας απαντήσω. Ας υποθέσουμε ότι ο όρος νοημοσύνη συνδέεται με την ικανότητα πολλών πραγμάτων, π.χ. αφηρημένη και λογική σκέψη, συνειδητότητα, δημιουργικότητα, μεταξύ άλλων. Τώρα, κάποια από αυτά φαίνεται να τα επιδεικνύουν τα μοντέρνα συστήματα τεχνητής νοημοσύνης. Εγώ θα έλεγα ότι, τουλάχιστον σε κάποιο βαθμό, εξομοιώνει πτυχές αυτού που ονομάζουμε νοημοσύνη.
 - Με εξιτάρει πάντως ότι ακόμα και επιστήμονες χρησιμοποιούν τέτοια ερωτήματα που προϋποθέτουν ότι έχουμε καλοσχηματισμένους ορισμούς για ζητήματα όπως η γνώση, η συνείδηση κ.λπ.

- Ξαναγυρνάμε πίσω στο προηγούμενο ερώτημα, τι ακριβώς σημαίνει κατανόηση της ΦΓ·
- Ένας τρόπος να δείξεις κάτι τέτοιο, το οποίο έχει τις βάσεις του στο τεστ του Τούρινκ για την Τεχνητή Νοημοσύνη, πάει ως εξής: αν έχεις κάποιον παρατηρητή σε μια συνομιλία μεταξύ ανθρώπου και μηχανής, αλλά δεν ξέρεις ποιος είναι ο άνθρωπος και ποιος είναι η μηχανή, τότε αν δεν μπορείς να τους ξεχωρίσεις από τη συμπεριφορά τους στη συνομιλία, τότε το μηχάνημα έχει περάσει το Τιούρινκ τεστ, έχει δείξει δείγματα ευφυίας

- Στοχαστικοί Παπαγάλοι (Bender et al. 2020):
χαρακτηρισμός των μεγάλων γλωσσικών μοντέλων.
Παπαγαλίζουν πράγματα τα οποία έχουν δει στα τεράστια δεδομένα που έχουν «δει», χωρίς καμία αναφορά στο νόημα
 - Τι είναι νόημα, παρόλα αυτά; Μπαίνουμε πάλι σε δύσκολες ερωτήσεις

- Για εμένα: αν εργασίες όπως η αυτόματη σύνοψη, η συνεπαγωγή φυσικής, γλώσσας, η αναγνώριση συναισθήματος, η λεξικολογική απόσταση μεταξύ λέξεων χρειάζονται κάποιου είδους κατανόηση για να εκτελεστούν σωστά, τότε, ναι, κάποιου είδους κατανόηση την προσομοιώνουν τα μεγάλα γλωσσικά μοντέλα

- Το ChatGPT είναι ένα τσάτμποτ TN. Στηρίζεται σε ένα μεγάλο γλωσσικό μοντέλο, το GPT-3.5, το οποίο έχει την ικανότητα να παράγει επαρκώς καλοσχηματισμένο κείμενο (φλυεντ τεξτ). Μοντέλα σαν το GPT-3.5 έχουν εκπαιδευτεί σε ένα τεράστιο αριθμό δεδομένων και το ChatGPT, που όπως είπαμε βασίζεται σε ένα τέτοιο μοντέλο, έχει επιπλέον προσαρμοστεί (fine-tuned) στο να παράγει λεπτομερείς απαντήσεις σε ερωτήματα που του θέτουν. Προσέξτε τον όρο «λεπτομερείς». Δεν προϋποθέτει την έννοια της ορθότητας.

- Μια εύλογη ανησυχία σε κοινωνικό-πολιτικό επίπεδο, ας πούμε έχει να κάνει με την αντικατάσταση ανθρωπίνου δυναμικού από συστήματα AI και τις ορδές ανέργων που προβλέπονται σε συγκεκριμένα εργασιακά πεδία – το ζήτημα απαιτεί αποφάσεις σε επίπεδο Πολιτείας. Μια ακόμη εύλογη ανησυχία έχει να κάνει με τη δυνατότητα χρήσης τέτοιας τεχνολογίας για διασπορά ψευδών ειδήσεων.

- Υπάρχουν σοβαρά οικολογικά ερωτήματα, μιας και το περιβαντολογικό αποτύπωμα που αφήνουν τέτοια μοντέλα για την εκπαίδευσή τους είναι τεράστιο. Πρέπει να δούμε ποιο είναι το περιβαντολογικό αντάλλαγμα για αυτά τα μοντέλα, να μπουν σοβαρά δηλαδή τέτοια ζητήματα στο τραπέζι. Για να λάβετε μια εικόνα του προβλήματος, σκεφτείτε ότι το μοντέλο GPT-3 χρειάστηκε 1287 μεγαβατώρες για την εκπαίδευσή του. Κάτι τέτοιο αντιστοιχεί στην ενέργεια που χρειάζεται ένα μέσο νοικοκυριό για 120 χρόνια.

- Singularity hypothesis (superintelligence)
 - Μας ξεπερνάνε σε ευφυΐα τα μηχανήματα κατά πολύ, σε κάποια φάση στα πλαίσια βελτιστοποίησης μιας εργασίας (ο Μπόστρομ αναφέρει την κατασκευή συνδετήρων), το μηχάνημα αποφασίζει ότι η καλύτερη δυνατή βελτιστοποίηση επιτυγχάνεται εξολοθρεύοντας όλο το ανθρώπινο είδος
 - Τι πιστεύω εγώ; Συμφωνώ με τους Dubhashi and Larrin (2017). Το γεγονός σενάριο αυτό αποτελεί λογική πιθανότητα. Όπως και η σύγκρουση και η καταστροφή της Γης από μετεωρίτη τα επόμενα πέντε χρόνια

- Ένα γλωσσικό μοντέλο είναι δυνατόν να δουλέψει σε πρακτικό επίπεδο παραβιάζοντας κάθε υπόθεση που ξέρουμε από τη Γλωσσολογία
- Οι μέθοδοι που έχουμε πλέον σε υπολογιστικό επίπεδο μπορούν να μας επιτρέψουν να τεστάρουμε υποθέσεις της Θεωρητικής Γλωσσολογίας και γενικότερα της Γνωσιακής Επιστήμης, οι οποίες παραμένουν αναπάντητες εδώ και πολλά χρόνια.

- Γλωσσική Κατάκτηση (χονδρικά: πώς καταφέρνει ένα παιδί να κατακτήσει πλήρως το σύστημα της γλώσσας)
 - τα ανθρώπινα όντα είναι γενετικά προκαθορισμένα για να κατακτήσουν τη γλώσσα. Αυτό απορρέει από την παρατήρηση ότι τα δεδομένα που λαμβάνει το παιδί είναι πολύ λίγα για να του επιτρέψουν να συνάγει, μέσω γενικότερων γνωσιακών διαδικασιών, ένα τόσο πολύπλοκο σύστημα, όπως η ανθρώπινη γλώσσα
- Υπάρχει η δυνατότητα πλέον, τουλάχιστον σε κάποιο βαθμό, να εξομοιώσουμε αυτές τις συνθήκες μάθησης και να δούμε αν πράγματι είναι δυνατό ή όχι ένας αλγόριθμος μάθησης να συνάγει ένα τέτοιο σύνθετο σύστημα.