



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΤΕΚΜΗΡΙΩΣΗ ΕΡΓΑΣΙΑΣ

ΜΑΘΗΜΑ
ΕΠΕΞΕΡΓΑΣΙΑ ΦΥΣΙΚΗΣ ΓΛΩΣΣΑΣ

ΘΕΜΑ ΕΡΓΑΣΙΑΣ: CHATBOT

ΕΠΙΜΕΛΕΙΑ ΕΡΓΑΣΙΑΣ
ΦΑΤΙΟΝ ΚΡΙΣΤΑΚΙ, Π20107

ΠΕΙΡΑΙΑΣ 2023

Σύντομη περιγραφή

Στην εργασία αυτή έχει υλοποιηθεί ένα chatbot για εστιατόριο που μπορεί να χειριστεί και να ανταπεξέλθει σε μερικές βασικές ερωτήσεις. Το περιβάλλον με το οποίο αλληλεπιδρά ο χρήστης είναι το terminal ενώ η εργασία έχει υλοποιηθεί σε python. Ο χρήστης αφού τρέξει το πρόγραμμα (εκτελεί το αρχείο main.py) μπορεί να ζητήσει διάφορες πληροφορίες σχετικά με το μενού, την τοποθεσία και άλλα. Το πρόγραμμα τερματίζει όταν ο χρήστης δεν χρειάζεται άλλες πληροφορίες και δίνει σαν input φράσεις όπως thank you, bye, good night και άλλα.

Αναλυτική περιγραφή

Αρχικά στο πρόγραμμα υπάρχουν κάποια δεδομένα από τα οποία αντλούμε τις πληροφορίες που μας χρειάζονται. Στο αρχείο data.py υπάρχει μια λίστα με κάθε κατηγορία ερωτήσεων. Κάθε κατηγορία είναι ένα αντικείμενο που έχει μια ετικέτα (πχ menu), κάποιες ενδεικτικές ερωτήσεις πάνω σε αυτό το θέμα καθώς και μερικές απαντήσεις.

Όταν τρέχουμε το πρόγραμμα συγκεντρώνουμε όλες τις πιθανές ερωτήσεις που μπορούν να γίνουν μαζί με την ετικέτα κάθε ερώτησης. Έτσι δημιουργούμε 2 πίνακες ίδιου μήκους με όλες τις ερωτήσεις και όλες τις αντίστοιχες ετικέτες. Πρέπει να σημειωθεί ότι κάθε ερώτηση από τον πίνακα με τις ερωτήσεις έχει φιλτραριστεί και έχουν αφαιρεθεί λέξεις που δεν έχουν μεγάλη σημασία (πχ άρθρα). Αυτό συμβαίνει για 2 λόγους. Πρώτον για να μην είναι πολύ περίπλοκο το μοντέλο μας με επιπλέον λέξεις μικρής έως μεσαίας σημασίας και δεύτερον για να είναι πιο ακριβής με τις απαντήσεις του ο αλγόριθμός μας.

Αφού συγκεντρώσουμε όλη την πληροφορία μας κάνουμε μια προεπεξεργασία δεδομένων. Στον πίνακα με τις ετικέτες αντιστοιχούμε κάθε κατηγορία με έναν αριθμό προκειμένου να μπορέσουμε να εκπαιδύσουμε το νευρωνικό μας δίκτυο.

Ύστερα ακολουθεί μια διαδικασία κατά την οποία αντιστοιχούμε σε όλες τις λέξεις των ερωτήσεων ένα βάρος αναλόγως την σημαντικότητά τους. Αυτό γίνεται ως εξής: Πρώτα δημιουργούμε έναν πίνακα μεγέθους αριθμός ερωτήσεων x αριθμός διακριτών λέξεων από όλες τις προτάσεις. Για κάθε πρόταση βάζουμε την τιμή 1 ή 0 σε κάθε στήλη ανάλογα με το αν η πρόταση αυτή περιέχει την λέξη αυτή ή όχι (χρήση βιβλιοθήκης CountVectorizer). Μετά από αυτό σε κάθε θέση του πίνακα αντιστοιχούμε ένα βάρος. Εάν η συγκεκριμένη λέξη εμφανίζεται σε πολλές προτάσεις τότε έχει μικρό βάρος διότι δεν μας δίνει χρήσιμη πληροφορία για να κάνουμε την ταξινόμησή μας. Αντιθέτως εάν υπάρχει κάποια χαρακτηριστική λέξη για κάποια κατηγορία που χρησιμοποιείται λίγες φορές τότε το βάρος αυτής της λέξης θα είναι μεγάλο. Το βάρος υπολογίζεται ως ο λογάριθμος του πλήθους των προτάσεων προς την συχνότητα των προτάσεων που περιέχουν την συγκεκριμένη λέξη (χρήση βιβλιοθήκης TfidfTransformer).

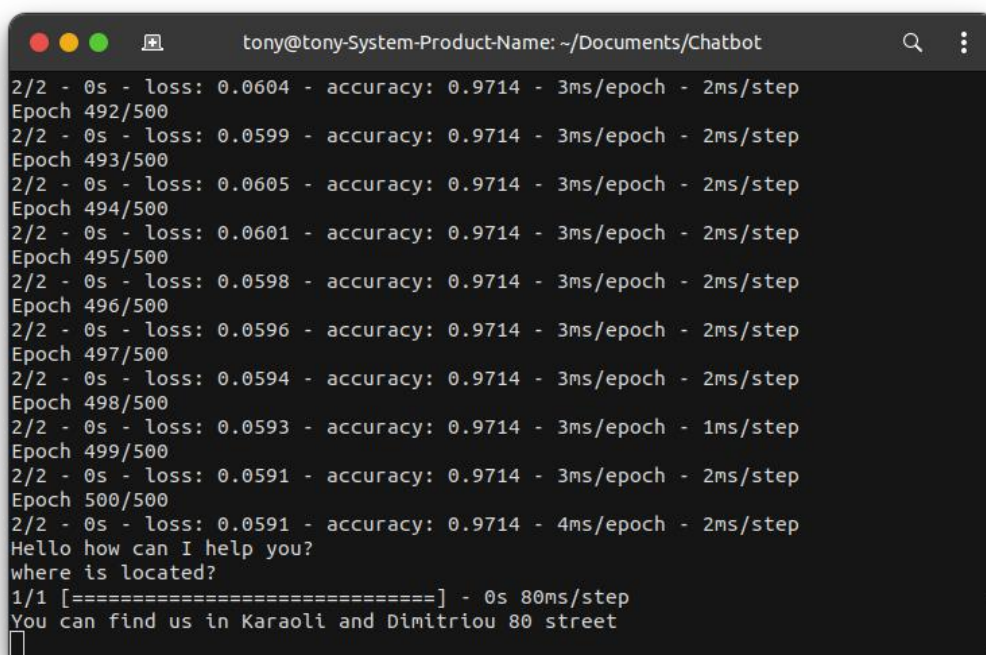
Αφού τελειώσουμε την παραπάνω διαδικασία, μετατρέπουμε την κάθε κατηγορία σε έναν πίνακα αριθμών 0 και 1 (μήκους όσο και οι κατηγορίες)

και εκπαιδεύουμε ένα πολυστρωματικό νευρωνικό δίκτυο με τον πίνακα των βαρέων και τον πίνακα των κατηγοριών.

Μετά το πέρας της εκπαίδευσης ξεκινάει ή αλληλεπίδραση με τον χρήστη αφού μέσω της κονσόλας ο χρήστης πληκτρολογεί την ερώτησή του. Η ερώτηση αναλύεται από όλα τα στάδια της προεπεξεργασίας δεδομένων που αναφέραμε και περνάει μέσα από το νευρωνικό μας δίκτυο το οποίο μας εμφανίζει την πιο πιθανή κατηγορία στην οποία ανήκει η ερώτηση. Ύστερα με βάση την ετικέτα αυτή το πρόγραμμα εκτυπώνει μια από τις απαντήσεις της κατηγορίας αυτής που υπάρχουν στη βάση μας. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να δοθεί σαν ερώτηση πρόταση από την κατηγορία "goodbye" όπου και τερματίζεται το πρόγραμμα.

Για το νευρωνικό μας δίκτυο χρησιμοποιούμε για τα hidden layers την relu ως activation function ($\max(x,0)$) καθώς και την softmax στο output layer η οποία ενδείκνυται για προβλήματα ταξινόμησης εξάγοντας την πιθανότητα κάθε κλάσης η συγκεκριμένη ερώτηση να ανήκει στην κατηγορία αυτή.

Screenshots εκτέλεσης



```
tony@tony-System-Product-Name: ~/Documents/Chatbot
2/2 - 0s - loss: 0.0604 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 492/500
2/2 - 0s - loss: 0.0599 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 493/500
2/2 - 0s - loss: 0.0605 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 494/500
2/2 - 0s - loss: 0.0601 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 495/500
2/2 - 0s - loss: 0.0598 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 496/500
2/2 - 0s - loss: 0.0596 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 497/500
2/2 - 0s - loss: 0.0594 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 498/500
2/2 - 0s - loss: 0.0593 - accuracy: 0.9714 - 3ms/epoch - 1ms/step
Epoch 499/500
2/2 - 0s - loss: 0.0591 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 500/500
2/2 - 0s - loss: 0.0591 - accuracy: 0.9714 - 4ms/epoch - 2ms/step
Hello how can I help you?
where is located?
1/1 [=====] - 0s 80ms/step
You can find us in Karaoli and Dimitriou 80 street
```

```
tony@tony-System-Product-Name: ~/Documents/Chatbot
Epoch 493/500
2/2 - 0s - loss: 0.0605 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 494/500
2/2 - 0s - loss: 0.0601 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 495/500
2/2 - 0s - loss: 0.0598 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 496/500
2/2 - 0s - loss: 0.0596 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 497/500
2/2 - 0s - loss: 0.0594 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 498/500
2/2 - 0s - loss: 0.0593 - accuracy: 0.9714 - 3ms/epoch - 1ms/step
Epoch 499/500
2/2 - 0s - loss: 0.0591 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 500/500
2/2 - 0s - loss: 0.0591 - accuracy: 0.9714 - 4ms/epoch - 2ms/step
Hello how can I help you?
where is located?
1/1 [=====] - 0s 80ms/step
You can find us in Karaoli and Dimitriou 80 street
what king of food do you have
1/1 [=====] - 0s 17ms/step
You can check the menu in our website www.example.gr/menu at your convinience
█
```

```
tony@tony-System-Product-Name: ~/Documents/Chatbot
2/2 - 0s - loss: 0.0601 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 495/500
2/2 - 0s - loss: 0.0598 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 496/500
2/2 - 0s - loss: 0.0596 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 497/500
2/2 - 0s - loss: 0.0594 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 498/500
2/2 - 0s - loss: 0.0593 - accuracy: 0.9714 - 3ms/epoch - 1ms/step
Epoch 499/500
2/2 - 0s - loss: 0.0591 - accuracy: 0.9714 - 3ms/epoch - 2ms/step
Epoch 500/500
2/2 - 0s - loss: 0.0591 - accuracy: 0.9714 - 4ms/epoch - 2ms/step
Hello how can I help you?
where is located?
1/1 [=====] - 0s 80ms/step
You can find us in Karaoli and Dimitriou 80 street
what king of food do you have
1/1 [=====] - 0s 17ms/step
You can check the menu in our website www.example.gr/menu at your convinience
thank you bye
1/1 [=====] - 0s 28ms/step
Goodbye!
tony@tony-System-Product-Name:~/Documents/Chatbot$ █
```