ΕΘΝΙΚΟ ΚΑΙ
ΚΑΠΟΔΙΣΤΡΙΑΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

Τεχνητή Νοημοσύνη **2**

3η Εργασία

Ιωάννης Δαλιάνης 1115201700027

Τμήμα Πληροφορικής και Τηλεπικοινωνιών, ΕΚΠΑ

Ιανουάριος 2021

Περιεχόμενα

Σελίδα

1 Προγραμματιστική Άσκηση

1.1 Σχόλια Υλοποίησης

Παρατίθεται το αρχείο ex3_RNN.ipynb στο οποίο υλοποιείται το bidirectional stacked RNN.
Αφαιρούνται από τα tweets πράγματα όπως tags, αριθμοί, πολλαπλοί συνεχόμενοι κενοί χαρακτήρες, yrls, κ.λ.π. Δε χρησιμοποιήθηκε stopwords removal διότι τα RNN δίνουν σημασία στις προτάσεις κατά την εκπαίδευση, οπότε αν αφαιρεθούν κάποιες κλασικές λέξεις μπορεί να χάνεται αρκετά το νόημα.
Χωρίζω τα δεδομένα σε δύο sets trainTWEETS και testTWEETS, τα οποία αποθηκεύονται για να διαβαστούν ως csv αρχεία από διαφορετικά paths από τη δομή trainTWEETS, testTWEETS, η οποία τα μετατράπει σε tuple.
Το LABEL αντιστοιχεί τα διαφορετικά labels του αρχικού dataset σε 0, 1, 2... Υπάρχει περίπτωση να αντιστοιχίσει τα labels που ήταν 1 στο αρχικό dataset με 0 και αυτά που ήταν 0 με 1. Αυτή η περίπτωση έχει προληφθεί στις αργότερα προβλέψεις.
Τα RNN παίρνουν σαν είσοδο μια ακολουθία λέξεων και για κάθε μία φτιάχνουν ένα hidden state μέσα στο δίκτυο χρησιμοποιώντας και τα hidden states των προηγούμενων λέξεων επαναληπτικά. Χρησιμοποιήθηκε η LSTM αρχιτεκτονική, η οποία σε αντίθεση με τα κλασικά RNN, δεν αντιμετωπίζει το πρόβλημα των vanishing gradients.
Χρησιμοποιώντας GloVe δίνουμε στο vocabulary μας προεκπαιδευμένους πίνακες με λέξεις παρόμοιας σημασίας χωρίς να χρειάζεται το μοντέλο μας να τους ανακαλύψει από μόνο του.
Η δομή data.BucketIterator ομαδοποιεί προτάσεις με παρόμοια lengths μαζί, ώστε να απαιτείται στη συνέχεια λιγότερο padding.

Μετά την εκπαίδευση, αξιολογούνται τα tweets του test αρχείου ένα προς ένα και τα αποτελέσματα εμφανίζονται σε ένα καινούριο pandas dataframe. Πάνω σε αυτό υπολογίζονται και εμφανίζονται οι μετρικές αξιολόγησης της ακρίβειας των προβλέψεων.
Σχολιασμός υπάρχει και στο notebbok αρχείο.
Παρακάτω παρατίθεται πίνακας με τα accuracy score για τις διάφορες αλληλουχίες παραμέτρων εκτέλεσης και εκπαίδευσης. Σε κάποιες εκτελέσεις χρησιμοποιήθηκε και early stopping με patience = 4.

1.2 Accuracy Score

- MVS MAX_VOCAB_SIZE
- BS BATCH_SIZE
- EMD EMBEDDING_DIM
- HD HIDDEN_DIM
- ACC ACCURACY
- NL N_LAYERS
- NE N_EPOCHS
- LR learning rate
- OVR overfit danger
- DROP DROPOUT
- EPST epoch that early stopping indicates if early stopping was used

Table 1.1: Scores

MVS	BS	EMD	HD	NL	NE	DROP	LR	ACC	OVR	EPST
25000	64	100	256	2	15	0.5	0.0001	0.8197	✗	✗
25000	64	100	300	2	15	0.5	0.0001	0.8199	✗	✗
25000	64	100	256	2	10	0.5	0.0001	0.8165	✗	✗
25000	64	100	300	2	11	0.5	0.0001	0.8174	✗	✗
25000	64	100	256	2	10	0.5	0.00001	0.7832	✗	✗
25000	64	100	300	2	11	0.5	0.001	0.8201	✓	✗
25000	64	100	256	3	10	0.5	0.0001	0.8168	✗	✗
25000	128	100	300	2	11	0.5	0.0001	0.8135	✗	✗
25000	64	100	256	3	10	0.6	0.0001	0.8137	✗	✗
30000	128	100	300	2	11	0.5	0.0001	0.8146	✗	✗
25000	128	100	256	3	10	0.7	0.0001	0.8048	✗	✗
30000	128	100	350	2	11	0.5	0.0001	0.8146	✗	✗
25000	128	100	256	3	11	0.3	0.0001	0.8181	✓	✗
25000	128	100	256	3	7	0.3	0.0001	0.8122	✗	✗
35000	256	100	300	2	11	0.5	0.0001	0.8109	✗	✗
25000	128	100	256	3	50	0.3	0.0001	0.8139	✗	14
35000	256	100	300	2	50	0.5	0.0001	0.8134	✗	24
35000	256	100	300	2	50	0.5	0.0001	0.8207	✗	25

1.3 Σχολιασμός Αποτελεσμάτων

Παρατηρούμε ότι τα score είναι παρόμοια. Σχετικά καλύτερο είναι είναι αυτό της τελευταίας γραμμής του πίνακα (0.8207). Παραδίδεται το μοντέλο με τις συγκεκριμένες παραμέτρους.
Το learning rate αν είναι μικρότερο από την τιμή 0.0001 βγάζει χειρότερα αποτελέσματα και χρειάζεται περισσότερο χρόνο εκπαίδευσης, καθιστώντας ανεπαίσθητες τις διαφορές του step size σε κάθε iteration. Αν είναι μεγαλύτερο, υπάρχει κίνδυνος να εμφανιστεί overfitting καθώς το μοντέλο προσπαθεί να βελτιώσει κατά πολύ τα score του μεταξύ διαδοχικών επαναλήψεων, άρα επικεντρώνεται εκτενώς στα μέχρι εκείνη τη στιγμή δεδομένα εκπαίδευσης.

2 Attention(Bonus)

Παρατίθεται το αρχείο ex3_RNN_Attention.ipynb στο οποίο υλοποιείται το bidirectional stacked RNN ακριβώς όπως και προηγουμένως, με την προσθήκη ενός Attention Layer. Στο συγκεκριμένο αρχείο για ένα πλήθος δεδομένων (κάτι παραπάνω από το 1/4) εκπαιδεύεται πρώτα το RNN χωρίς Attention με εμφάνιση αποτελεσμάτων και στη συνέχει εκπαιδεύεται το RNN με Attention εμφανίζοντας και τα αντίστοιχα αποτελέσματα.
Οι παράμετροι εκπαίδευσης είναι οι ίδιες με του καλύτερου μοντέλου όπως περιγράφηκε στο 1.3.
Χωρίς τη χρήση Attention Layer το Early Stopping μας σταματάει στα 32 epochs έχοντας Training Accuracy 0.8177, Validation Accuracy 0.8051, Test Loss: 0.421 και Test Accuracy 80.52%.
Με χρήση Attention Layer το Early Stopping μας σταματάει στα 26 epochs έχοντας Training Accuracy 0.8108, Validation Accuracy 0.8025, Test Loss: 0.431 και Test Accuracy 80.25%.