

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ "ΠΡΟΗΓΜΕΝΑ ΣΥΣΤΗΜΑΤΑ
ΠΛΗΡΟΦΟΡΙΚΗΣ"



Εργασία Μαθήματος ***Προηγμένη Βαθιά Μάθηση με Python***

<i>Αριθμός εργασίας – Τίτλος εργασίας</i>	<i>Απαλλακτική</i>
Όνομα φοιτητή	Καλλίγερος Αναστάσιος
Αρ. Μητρώου	Mrsp2314
Ημερομηνία παράδοσης	28/9/2204

Εκφώνηση εργασίας



Advanced Deep Learning with Python
Spring Semester 2022
Course Assignment
Neural Machine Translation at the Word Level

- I. Provide an extensive theoretical description of the RNN-based Machine Translation approach that was adopted throughout the computational project “Word Level Machine Translation” that was presented in class. You should focus on describing the various components of the Encoder – Decoder network architecture both mathematically and graphically.
- II. Provide an extensive theoretical description of the Attention Mechanisms proposed for the problem of Neural Machine Translation presented by the following papers (both papers are available on gunet):
 - a. **“Neural Machine Translation by jointly learning to Align and Translate”** by Bahdanau et al. and
 - b. **“Effective Approaches to Attention-based Neural Machine Translation”** by Luong et al.Identify the major difference between the aforementioned approaches and the one presented in class. Elaborate on the major differences between the two attention mechanisms provided by Bahdanau et al. and Luong et al.
- III. Extend the existing code project by implementing both attention mechanisms.
- IV. Extend the existing code project by implemented the transformer model presented in the paper **“Attention is all you need”** by Vaswani et al.
- V. Your code should provide a comparative evaluation of the various models in terms of the testing accuracy and the quality of machine translation.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

1 Εισαγωγή.....	4
2 Επίδειξη της λύσης	4



Θεωρητική Περιγραφή της Μηχανικής Μετάφρασης με Βάση το RNN (Recurrent Neural Network)	4
Αρχιτεκτονική Κωδικοποιητή-Αποκωδικοποιητή (Encoder-Decoder)....	5
3 Βιβλιογραφικές Πηγές	18



1 Εισαγωγή

Η μηχανική μετάφραση αποτελεί ένα από τα σημαντικότερα πεδία έρευνας στην επεξεργασία φυσικής γλώσσας (NLP), με στόχο τη μετάφραση κειμένων από μία γλώσσα σε μία άλλη μέσω υπολογιστικών μεθόδων. Μια από τις πιο διαδεδομένες προσεγγίσεις για την επίτευξη αυτού του στόχου είναι η χρήση Αναδρομικών Νευρωνικών Δικτύων (RNNs) σε συνδυασμό με την αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή (encoder-decoder). Αυτή η αρχιτεκτονική επιτρέπει την αποτελεσματική μοντελοποίηση της αλληλουχίας λέξεων σε διαφορετικές γλώσσες και χρησιμοποιείται ευρέως για την επίλυση προβλημάτων μηχανικής μετάφρασης.

Στην παρούσα εργασία, θα γίνει μια εκτενής ανάλυση της μηχανικής μετάφρασης με βάση την προσέγγιση του RNN, όπως αυτή υλοποιείται στο υπολογιστικό έργο “Word Level Machine Translation” που παρουσιάστηκε στην τάξη. Η ανάλυση θα εστιάσει στην περιγραφή των διαφόρων στοιχείων της αρχιτεκτονικής του δικτύου κωδικοποιητή-αποκωδικοποιητή, τόσο σε θεωρητικό επίπεδο, με μαθηματική προσέγγιση, όσο και σε πρακτικό επίπεδο, με αναφορά στον συγκεκριμένο κώδικα.

2 Επίδειξη της λύσης

I.

Θεωρητική Περιγραφή της Μηχανικής Μετάφρασης με Βάση το RNN (Recurrent Neural Network)

Η μηχανική μετάφραση είναι ένα υποπεδίο της επεξεργασίας φυσικής γλώσσας (NLP) που έχει ως στόχο τη μετάφραση κειμένου από μία γλώσσα σε μία άλλη με τη βοήθεια υπολογιστικών μοντέλων. Η προσέγγιση που βασίζεται σε Αναδρομικά Νευρωνικά Δίκτυα (RNNs) για μηχανική μετάφραση έχει υιοθετηθεί ευρέως, ειδικά μέσω της αρχιτεκτονικής κωδικοποιητή-αποκωδικοποιητή (encoder-decoder). Αυτή η προσέγγιση



χρησιμοποιείται για να μοντελοποιήσει τη σχέση μεταξύ της πηγής και της στόχου γλώσσας.

Αρχιτεκτονική Κωδικοποιητή-Αποκωδικοποιητή (Encoder-Decoder)

Η αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή αποτελείται από δύο κύρια συστατικά:

- **Κωδικοποιητής (Encoder):** Επεξεργάζεται την ακολουθία εισόδου (π.χ., πρόταση στη γλώσσα προέλευσης) και τη μετατρέπει σε μια διανυσματική αναπαράσταση, συνήθως γνωστή ως "context vector" ή "hidden state".
- **Αποκωδικοποιητής (Decoder):** Χρησιμοποιεί την παραγόμενη διανυσματική αναπαράσταση από τον κωδικοποιητή για να παράγει την ακολουθία εξόδου (π.χ., πρόταση στη γλώσσα στόχο).

Κωδικοποιητής (Encoder)

Ο κωδικοποιητής αποτελείται από ένα ή περισσότερα στρώματα RNN, τα οποία λαμβάνουν ως είσοδο την ακολουθία λέξεων από την πηγή γλώσσα. Η κάθε λέξη της ακολουθίας μετατρέπεται σε ένα ενσωματωμένο διάνυσμα (embedding vector) και στη συνέχεια εισάγεται σε ένα RNN cell, όπως το LSTM (Long Short-Term Memory) ή GRU (Gated Recurrent Unit). Η έξοδος του κάθε RNN cell είναι ένα διάνυσμα καταστάσεων που περικλείει πληροφορίες από όλες τις προηγούμενες λέξεις της ακολουθίας.

Αποκωδικοποιητής (Decoder)

Ο αποκωδικοποιητής χρησιμοποιεί το διάνυσμα κατάστασης από τον κωδικοποιητή για να παράγει την έξοδο. Το πρώτο βήμα του αποκωδικοποιητή είναι να λάβει το κρυφό διάνυσμα του κωδικοποιητή ως αρχική κατάσταση και να ξεκινήσει να παράγει λέξεις για την πρόταση στη γλώσσα στόχο.

Πιο συγκεκριμένα όσον αφορά το project “**Word Level Machine Translation**”:



Αναπαράσταση Δεδομένων και Embeddings

Τα δεδομένα εισόδου και τα δεδομένα εξόδου αναπαρίστανται μέσω διανυσμάτων λέξεων (word embeddings). Τα embeddings είναι πίνακες (matrices) που μετατρέπουν τις λέξεις σε πυκνά διανύσματα χαμηλών διαστάσεων, τα οποία περιέχουν σημασιολογική πληροφορία.

Χαρακτηριστικό παράδειγμα: EMBEDDING_DIM

= 256

encoder =

Encoder(data_preparation.english_vocabulary_size+1, EMBEDDING_DIM,

data_preparation.english_maxlen, ENCODER_DIM)

decoder =

Decoder(data_preparation.french_vocabulary_size+1, EMBEDDING_DIM,

data_preparation.french_maxlen, DECODER_DIM)

Στον κώδικα, τα embeddings έχουν διαστάσεις 256 και χρησιμοποιούνται τόσο από τον κωδικοποιητή όσο και από τον αποκωδικοποιητή.

Κωδικοποιητής (Encoder)

Ο κωδικοποιητής είναι ένα RNN (Recurrent Neural Network) το οποίο μπορεί να είναι υλοποιημένο είτε με GRU (Gated Recurrent Unit) είτε με LSTM (Long Short-Term Memory). Το RNN επεξεργάζεται τις λέξεις της πρότασης μία προς μία, ενημερώνοντας την εσωτερική του κατάσταση (hidden state) σε κάθε βήμα.

Χαρακτηριστικό παράδειγμα:

encoder_out, encoder_state = encoder(encoder_in, encoder_state)



- **Είσοδος (encoder_in):** Η είσοδος στον κωδικοποιητή είναι η ακολουθία των διανυσμάτων λέξεων της πρότασης στα Αγγλικά.
- **Κατάσταση Κωδικοποιητή (encoder_state):** Η κρυφή κατάσταση αρχικοποιείται και ενημερώνεται σε κάθε χρονικό βήμα καθώς η πρόταση επεξεργάζεται.
- **Έξοδος (encoder_out):** Η έξοδος του κωδικοποιητή είναι η τελική κρυφή κατάσταση (ή σειρά καταστάσεων), η οποία συνοψίζει την πληροφορία της πρότασης εισόδου.

Αποκωδικοποιητής (Decoder)

Ο αποκωδικοποιητής είναι επίσης ένα RNN, το οποίο χρησιμοποιεί την κατάσταση του κωδικοποιητή για να παράγει την πρόταση στη γλώσσα στόχο. Η είσοδος του αποκωδικοποιητή είναι είτε η πραγματική λέξη της προηγούμενης χρονικής στιγμής (Teacher Forcing) είτε η λέξη που προέβλεψε ο ίδιος ο αποκωδικοποιητής στην προηγούμενη χρονική στιγμή.

Χαρακτηριστικό παράδειγμα:

```
decoder_pred, decoder_state = decoder(decoder_in, decoder_state)
```

- **Είσοδος (decoder_in):** Η είσοδος στον αποκωδικοποιητή είναι είτε η πρόβλεψη της προηγούμενης λέξης είτε η πραγματική λέξη της προηγούμενης χρονικής στιγμής.
- **Κατάσταση Αποκωδικοποιητή (decoder_state):** Αρχικά, αυτή είναι η τελική κατάσταση του κωδικοποιητή, αλλά ενημερώνεται σε κάθε χρονικό βήμα.
- **Έξοδος (decoder_pred):** Η έξοδος του αποκωδικοποιητή είναι η πρόβλεψη της επόμενης λέξης.



Teacher Forcing

Η τεχνική του Teacher Forcing χρησιμοποιείται κατά την εκπαίδευση, όπου η είσοδος στον αποκωδικοποιητή δεν είναι η λέξη που προέβλεψε το μοντέλο, αλλά η πραγματική λέξη της προηγούμενης χρονικής στιγμής. Αυτό βοηθά το μοντέλο να συγκλίνει ταχύτερα.

Χαρακτηριστικό παράδειγμα:

```
decoder_pred, decoder_state = decoder(decoder_in, decoder_state)
```

Λειτουργία Απώλειας (Loss Function)

Η λειτουργία απώλειας που χρησιμοποιείται είναι η `SparseCategoricalCrossentropy`. Σημαντικό είναι το γεγονός ότι η λειτουργία αυτή αγνοεί τα "pad" tokens, ώστε να αξιολογείται σωστά η πρόοδος του μοντέλου.

Χαρακτηριστικό παράδειγμα:

```
def loss_fn(ytrue, ypred):  
    scce = tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True)  
    mask = tf.math.logical_not(tf.math.equal(ytrue, 0))  
    mask = tf.cast(mask, dtype=tf.int64)  
    loss = scce(ytrue, ypred, sample_weight=mask)  
    return loss
```

Εκπαίδευση του Μοντέλου

Το μοντέλο εκπαιδεύεται για έναν αριθμό επαναλήψεων (epochs). Σε κάθε εποχή, το μοντέλο ενημερώνεται με βάση τα δεδομένα εκπαίδευσης, και αξιολογείται με τη χρήση του BLEU score.

Χαρακτηριστικό παράδειγμα:



```
eval_scores, losses = train_model(EPOCHS_NUMBER, DELTA_EPOCHS,  
encoder, decoder)
```

Γραφική Αναπαράσταση και Αξιολόγηση

Κατά την εκπαίδευση, παρακολουθούνται δύο κύριες μετρικές: το BLEU Score, το οποίο αξιολογεί την ποιότητα της μετάφρασης, και η λειτουργία απώλειας, που δείχνει την απόκλιση του μοντέλου από το ιδανικό αποτέλεσμα.

Χαρακτηριστικό παράδειγμα:

```
plt.plot(eval_scores) plt.title('Model Accuracy in  
terms of the BLEU Score.') plt.ylabel('BLEU Score')  
plt.xlabel('epoch') plt.legend(['BLEU Score'],  
loc='lower right') plt.grid() plt.show() plt.plot(losses)  
plt.title('Model Accuracy in terms of the Loss.')  
plt.ylabel('Loss') plt.xlabel('epoch')  
plt.legend(['LOSS'], loc='upper right') plt.grid()  
plt.show()
```

Συμπερασματικά η προσέγγιση RNN με κωδικοποιητή-αποκωδικοποιητή είναι ένα ισχυρό εργαλείο για τη μηχανική μετάφραση. Η χρήση τεχνικών όπως **to Teacher Forcing** και η βελτιστοποίηση με βάση την απώλεια **CrossEntropy** επιτρέπουν στο μοντέλο να εκπαιδεύεται αποτελεσματικά. Με την κατάλληλη ρύθμιση των υπερπαραμέτρων και την επαρκή εκπαίδευση, το μοντέλο μπορεί να επιτύχει υψηλής ποιότητας μεταφράσεις.



II.

Οι μηχανισμοί προσοχής έχουν αναδειχθεί ως σημαντική καινοτομία στη νευρωνική μηχανική μετάφραση (Neural Machine Translation - NMT), βελτιώνοντας σημαντικά την ακρίβεια και την αποτελεσματικότητα των συστημάτων μετάφρασης. Οι δύο εργασίες που αναφέρονται, «Neural Machine Translation by jointly learning to Align and Translate» του Bahdanau et al. και «Effective Approaches to Attention-based Neural Machine Translation» των Luong et al., περιγράφουν διαφορετικές προσεγγίσεις για την εφαρμογή αυτών των μηχανισμών προσοχής. Ακολουθεί μια εκτενής θεωρητική περιγραφή των προτεινόμενων μηχανισμών προσοχής στις δύο αυτές εργασίες.

“Neural Machine Translation by jointly learning to Align and Translate” by Bahdanau et al.

Η εργασία των Bahdanau et al. εισήγαγε για πρώτη φορά την ιδέα του μηχανισμού προσοχής στη νευρωνική μηχανική μετάφραση. Πριν από αυτήν την εργασία, τα περισσότερα NMT μοντέλα βασίζονταν σε μοντέλα ακολουθίας-σε-ακολουθία (sequence-to-sequence ή seq2seq) με κωδικοποιητή-αποκωδικοποιητή (encoder-decoder), όπου ο κωδικοποιητής (encoder) μετατρέπει την εισερχόμενη ακολουθία σε ένα σταθερού μεγέθους διάνυσμα (context vector), το οποίο στη συνέχεια χρησιμοποιείται από τον αποκωδικοποιητή (decoder) για την παραγωγή της εξόδου. Το πρόβλημα με αυτήν την προσέγγιση ήταν ότι το διάνυσμα περιβάλλοντος είχε περιορισμένη ικανότητα να συλλάβει όλες τις πληροφορίες από την είσοδο, ιδιαίτερα για μεγάλες προτάσεις.

Προτεινόμενος Μηχανισμός Προσοχής:

Οι Bahdanau et al. πρότειναν έναν μηχανισμό ευθυγράμμισης (alignment) που λειτουργεί ως προσοχή (attention) και επιτρέπει στο μοντέλο να «εστιάζει» σε διαφορετικά μέρη της εισόδου σε κάθε βήμα της διαδικασίας παραγωγής της εξόδου.



Ευθυγράμμιση (Alignment): Σε κάθε βήμα της διαδικασίας αποκωδικοποίησης, το μοντέλο υπολογίζει ένα σύνολο βαρών προσοχής (attention weights) που αναπαριστούν πόση προσοχή πρέπει να δίνεται σε κάθε λέξη της εισόδου για να παραχθεί η τρέχουσα λέξη της εξόδου. Αυτά τα βάρη προκύπτουν μέσω μιας συνάρτησης σκόρ (score function), που μετρά την συνάφεια (relevance) ανάμεσα στην τρέχουσα κατάσταση του αποκωδικοποιητή (decoder hidden state) και τα κρυφά καταστάσεις του κωδικοποιητή (encoder hidden states).

Υπολογισμός των βαρών προσοχής:

$$e_{ij} = f(s_{i-1}, h_j)$$

όπου το e_{ij} είναι το score που αντιστοιχεί στην j-οστή λέξη της εισόδου και την i-οστή λέξη της εξόδου, s_{i-1} είναι η κρυφή κατάσταση του αποκωδικοποιητή στο προηγούμενο βήμα, και h_j είναι η κρυφή κατάσταση του κωδικοποιητή για την j-οστή λέξη της εισόδου.

Συνάρτηση προσοχής (Attention function):

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}$$

Τα βάρη προσοχής α_{ij} κανονικοποιούνται χρησιμοποιώντας την εκθετική συνάρτηση softmax ώστε να παράγουν πιθανότητες.

Πλαίσιο προσοχής (Context Vector): Το τελικό διάνυσμα πλαισίου (context vector) για κάθε βήμα αποκωδικοποίησης είναι ο σταθμισμένος μέσος όρος των κρυφών καταστάσεων του κωδικοποιητή:



Αυτό το διάνυσμα χρησιμοποιείται μαζί με την προηγούμενη κρυφή κατάσταση του αποκωδικοποιητή για την πρόβλεψη της τρέχουσας λέξης της εξόδου.

Effective Approaches to Attention-based Neural Machine Translation (Luong et al., 2015)

Η εργασία των Luong et al. (2015) επεκτείνει και βελτιώνει την αρχική προσέγγιση του Bahdanau et al., προτείνοντας διαφορετικές παραλλαγές του μηχανισμού προσοχής και εξετάζοντας τις επιδόσεις τους σε διάφορες γλώσσες και ζεύγη γλωσσών.

Παραλλαγές Μηχανισμών Προσοχής:

1. Global vs Local Attention:

- **Global Attention:** Αυτή η προσέγγιση είναι παρόμοια με αυτήν του Bahdanau et al., όπου η προσοχή υπολογίζεται χρησιμοποιώντας όλες τις κρυφές καταστάσεις του κωδικοποιητή για να δημιουργηθεί το διάνυσμα πλαισίου.
 - **Local Attention:** Αυτή η παραλλαγή εισάγει την ιδέα της «τοπικής προσοχής», όπου η προσοχή εστιάζει μόνο σε ένα υποσύνολο των κρυφών καταστάσεων του κωδικοποιητή. Αυτό το υποσύνολο επιλέγεται γύρω από (predictive position), $c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$ μια προκαθορισμένη θέση μειώνοντας τον υπολογιστικό φόρτο και βελτιώνοντας την αποδοτικότητα.
2. **Σκορ προσοχής (Attention Score):** Οι Luong et al. πρότειναν διαφορετικές συναρτήσεις σκόρ για τον υπολογισμό των scores e_{ij} , όπως:



Dot:
$$e_{ij} = s_{i-1}^T h_j$$

General:
$$e_{ij} = s_{i-1}^T W_a h_j$$

όπου W_a είναι μία παράμετρος που πρέπει να μάθει το μοντέλο.

Concatenate:
$$e_{ij} = v_a^T \tanh(W_a[s_{i-1}; h_j])$$

όπου W_a και v_a είναι παράμετροι του μοντέλου, και $[\cdot; \cdot]$ αναπαριστά την συνένωση (concatenation) των διανυσμάτων.

3. Στρατηγικές προσοχής (Attention Strategies):

- **Hard Attention:** Μία μορφή προσοχής όπου γίνεται ξεκάθαρη επιλογή της θέσης εισόδου που θα ληφθεί υπόψη, κάτι που καθιστά την προσοχή μη διαφορίσιμη.
- **Soft Attention:** Η παραδοσιακή μορφή όπου υπολογίζονται και συνδυάζονται όλοι οι πιθανοί στόχοι με βάρη που κανονικοποιούνται.

Συμπερασματικά οι δύο προσεγγίσεις εισάγουν καινοτομίες στην νευρωνική μηχανική μετάφραση μέσω της εφαρμογής των μηχανισμών προσοχής. Η εργασία του Bahdanau et al. είναι πρωτοποριακή, καθώς εισήγαγε τον μηχανισμό προσοχής που επιτρέπει στο μοντέλο να «δει» ολόκληρη την εισαγωγική ακολουθία σε κάθε βήμα της μετάφρασης. Η εργασία των Luong et al. βελτιώνει και επεκτείνει αυτήν την ιδέα, προτείνοντας διαφορετικές μορφές προσοχής (global και local) και νέες συναρτήσεις σκωρ που προσφέρουν μεγαλύτερη ευελιξία και βελτίωση της απόδοσης του συστήματος.



Συνολικά, οι μηχανισμοί προσοχής επέτρεψαν στα μοντέλα NMT να διαχειρίζονται καλύτερα τις μεγάλες προτάσεις, να διατηρούν την συνάφεια της μετάφρασης και να επιτυγχάνουν καλύτερα αποτελέσματα σε σύγκριση με τα παραδοσιακά συστήματα μηχανικής μετάφρασης.

Για να κατανοήσουμε τις διαφορές μεταξύ των προσεγγίσεων προσοχής που προτείνονται από τους Bahdanau et al. και τους Luong et al., καθώς και της προσέγγισης που παρουσιάστηκε στην τάξη, πρέπει να εξετάσουμε τη λειτουργία και τη φιλοσοφία πίσω από αυτές τις μεθόδους. Παρακάτω παρουσιάζονται οι σημαντικότερες διαφορές και οι βασικές αρχές που τις διακρίνουν.

1. Βασική διαφορά με την προσέγγιση που παρουσιάστηκε στην τάξη

Η προσέγγιση που παρουσιάστηκε στην τάξη βασίζεται στο παραδοσιακό sequence-to-sequence μοντέλο με ένα απλό κωδικοποιητή-αποκωδικοποιητή χωρίς μηχανισμό προσοχής, το οποίο ήταν η κύρια μέθοδος πριν την εισαγωγή του μηχανισμού προσοχής από τον Bahdanau et al. Σε αυτήν την παραδοσιακή προσέγγιση:

- **Context Vector:** Ολόκληρη η ακολουθία εισόδου συμπίεζεται σε ένα σταθερού μεγέθους διάνυσμα (context vector) από τον κωδικοποιητή. Αυτό το διάνυσμα περιβάλλοντος είναι το μόνο που μεταφέρεται στον αποκωδικοποιητή για τη δημιουργία της εξόδου.
- **Κύριο πρόβλημα:** Για μεγάλες ακολουθίες, αυτό το διάνυσμα δεν μπορεί να καταγράψει όλες τις λεπτομέρειες της εισόδου, οδηγώντας σε φτωχότερη απόδοση στις μεταφράσεις.

Οι μηχανισμοί προσοχής, όπως αυτοί που προτάθηκαν από τους Bahdanau et al. και τους Luong et al., αντιμετωπίζουν αυτό το πρόβλημα δίνοντας στο



μοντέλο τη δυνατότητα να «εστιάζει» σε διαφορετικά μέρη της ακολουθίας εισόδου κατά τη διάρκεια της διαδικασίας αποκωδικοποίησης.

2. Σημαντικότερες διαφορές μεταξύ των προσεγγίσεων των Bahdanau et al. και των Luong et al.

1. Μηχανισμός Ευθυγράμμισης:

- Η προσέγγιση αυτή εισήγαγε την έννοια της «ευθυγράμμισης» (alignment), όπου η προσοχή επιτρέπει στο μοντέλο να συνδυάζει πληροφορίες από συγκεκριμένα σημεία της εισόδου όταν παράγει κάθε λέξη της εξόδου.
- Σε κάθε βήμα της διαδικασίας αποκωδικοποίησης, υπολογίζεται ένα διάνυσμα πλαισίου (context vector) που είναι ο σταθμισμένος μέσος όρος όλων των κρυφών καταστάσεων του κωδικοποιητή.

2. Συνάρτηση Ευθυγράμμισης (Alignment Function):

- Χρησιμοποιείται μια συνάρτηση σκορ (score function) για τον υπολογισμό της συνάφειας μεταξύ της τρέχουσας κρυφής κατάστασης του αποκωδικοποιητή και κάθε κρυφής κατάστασης του κωδικοποιητή.
- Η softmax εφαρμόζεται στα αποτελέσματα της συνάρτησης σκορ για να παράγει τα βάρη προσοχής.

3. Ευκαμψία:

- Το μοντέλο μπορεί να επιλέξει πόση προσοχή θα δώσει σε κάθε λέξη της εισόδου ανάλογα με το πόσο σχετική είναι με την τρέχουσα λέξη εξόδου, κάτι που προσφέρει μεγάλη ευελιξία στο σύστημα.



β. Luong et al. (2015): Διάφορες Προσεγγίσεις Προσοχής (Various Attention Approaches)

1. Global vs Local Attention:

- **Global Attention:** Αυτή η προσέγγιση είναι παρόμοια με αυτήν των Bahdanau et al., καθώς χρησιμοποιεί όλες τις κρυφές καταστάσεις του κωδικοποιητή για να υπολογίσει το διάνυσμα πλαισίου σε κάθε βήμα αποκωδικοποίησης.
- **Local Attention:** Εισάγεται μια νέα έννοια όπου η προσοχή εστιάζει σε ένα υποσύνολο των κρυφών καταστάσεων του κωδικοποιητή. Αυτό επιτυγχάνεται προβλέποντας τη θέση προσοχής και επικεντρώνοντας την προσοχή γύρω από αυτήν τη θέση. Αυτό μπορεί να βελτιώσει την αποδοτικότητα και να μειώσει τον υπολογιστικό φόρτο.

2. Σκορ Προσοχής (Attention Score):

- Οι Luong et al. πρότειναν διάφορες συναρτήσεις για τον υπολογισμό του score προσοχής:
 - **Dot:** Το απλούστερο score που χρησιμοποιεί το εσωτερικό γινόμενο (dot product) μεταξύ της κρυφής κατάστασης του αποκωδικοποιητή και κάθε κρυφής κατάστασης του κωδικοποιητή.
 - **General:** Μια γενικότερη μορφή όπου το score υπολογίζεται με το εσωτερικό γινόμενο αφού πρώτα πολλαπλασιαστεί η κρυφή κατάσταση του κωδικοποιητή με μία μητρώα παράμετρο.
 - **Concatenate:** Το score υπολογίζεται συνδυάζοντας και τις δύο κρυφές καταστάσεις και εφαρμόζοντας μια μη γραμμική συνάρτηση.



3. Ευκαμψία και Απόδοση:

- Οι Luong et al. διερεύνησαν την αποδοτικότητα αυτών των διαφορετικών μορφών προσοχής και έδειξαν ότι η επιλογή του κατάλληλου τύπου προσοχής μπορεί να βελτιώσει τις επιδόσεις ανάλογα με το μέγεθος της ακολουθίας εισόδου και τη γλώσσα.

Συνοπτική Σύγκριση των Bahdanau et al. και Luong et al.

- **Bahdanau et al.:** Η εστίαση είναι στη δυναμική ευθυγράμμιση της εισόδου με την έξοδο, με τον μηχανισμό προσοχής να αποφασίζει σε κάθε βήμα αποκωδικοποίησης πόση προσοχή να δοθεί σε κάθε τμήμα της εισόδου. Η προσέγγιση αυτή είναι πιο σύνθετη, καθώς χρησιμοποιεί έναν μηχανισμό ευθυγράμμισης για να καθορίσει τη συνάφεια μεταξύ των εισόδων και των εξόδων.
- **Luong et al.:** Η προσέγγιση αυτή είναι πιο ευέλικτη, προσφέροντας πολλαπλές παραλλαγές της προσοχής (global vs. local) και διάφορες συναρτήσεις για τον υπολογισμό του score προσοχής. Η «Local Attention» είναι μια πιο υπολογιστικά αποδοτική εκδοχή που προσπαθεί να περιορίσει την περιοχή της προσοχής, μειώνοντας το συνολικό φόρτο επεξεργασίας.

Συμπερασματικά η κύρια διαφορά μεταξύ των δύο προσεγγίσεων είναι ο τρόπος με τον οποίο διαχειρίζονται και υπολογίζουν την προσοχή. Οι **Bahdanau et al.** εισήγαγαν την προσοχή ως ένα δυναμικό σύστημα ευθυγράμμισης, ενώ οι **Luong et al.** παρουσίασαν μια πιο γενική προσέγγιση με πολλαπλές επιλογές και βελτιστοποιήσεις. Κάθε προσέγγιση έχει τις δικές της εφαρμογές και πλεονεκτήματα, ανάλογα με τις απαιτήσεις του συστήματος και τα δεδομένα που χρησιμοποιούνται.



3 Βιβλιογραφικές Πηγές

- “Neural Machine Translation by jointly learning to Align and Translate”
by Bahdanau et al.
- “Effective Approaches to Attention-based Neural Machine Translation”
by Luong et al.