



Фигура 1: Схема на рекурентната невронна мрежа, реализираща невронен машинен превод. Encoder-Decoder с механизъм за внимание.



## Кратко описание на архитектурата.

Моделът използва encoder-decoder с механизъм за внимание.  $LSTM_E$  представя encoder, а  $LSTM_D$  - decoder.  $LSTM_E$  е двупосотна рекурентна невронна мрежа от тип LSTM. Като първоначално състояние се задава нулевият вектор/матрица (тензор).  $LSTM_D$  е рекурентна еднопос. невронна мрежа от тип LSTM. Броят на скритите слоеве на  $LSTM_D$  е два пъти броят на скритите слоеве на  $LSTM_E$ . Като първоначално състояние на  $LSTM_D$  се задава ~~то~~ последното състояние на  $LSTM_E$ . След получаване на изходните вектори на двете невронни мрежи ( $h_E$  и  $h_D^x$  от фиг. 1) се прилага Dropout върху тях. ( $\bar{h}_E = \text{dropout}(h_E)$ ,  $\bar{h}_D = \text{dropout}(h_D)$ ). Върху векторите  $\bar{h}_E$  и  $\bar{h}_D$  (от фиг. 1) се прилага адитивният механизъм за внимание [2]. След получаване на вектора на внимание, той се припелва към изходния вектор на  $LSTM_D$  след Dropout и новополученият вектор се подава на перцептрона Projection, тъйто броят на неврони в изходния слой е равен на броят на думите в речника на целевия език. След това се прилага Softmax и се избира най-вероятната дума.

## Обучение на модела

При входа на декодера се подава съответната дума от целевото изречение, а не най-вероятната дума, генерирана на предишната стъпка от декодера.

За обучението на модела са използвани предоставе-



ните програми. (спускане по градиента с партиден стохастичен градиент).

Размерът на партидата е 20, ~~както~~<sup>и 24.</sup> е при обучение на модела, а епохите са 5.

Резултати:

batch-size: размери на партидата

embedding-size: размерност на влизането

hidden-size: брой неврони в скритите слоеве на енкодера и декодера.

num\_layers: брой на слоевете на декодера (при енкодера са два пъти по-малко)

dropout:

batch-size	24	20
embedding-size:	1024	1024
hidden-size:	256	256
num_layers	2	4
dropout	0.2	0.2
Време за обучение	5923 секунди	10445 секунди
перплексия върху тестовия корпус	4.8640057	5.0872418
анско търсене	27.87581	33.91718
анско търсене (при link се записва сбс следващата най-вероятна дума)	29.45925	35.02788
Beam search ( $\beta=2$ )	30.26344	35.93028
Beam search ( $\beta=5$ )	31.61577	36.93319
Beam search ( $\beta=10$ )	32.11427	37.37265

BLEU  
оценка  
върху  
тестовия  
корпус

Източници: [1] Neural Machine Translation by Jointly Learning to Align and Translate, Bahdanau, Bengio  
[2] Massive Exploration of Neural Machine Architectures, Britz, Goldie, Luong, Le