

Project team: Omar Alhadlaq (hadlaq@) Saud Alsaif (salsaif@) and Yaya Khoja (ykhoja@)

Project Motivation

- Automating the generation of useful headlines for journalists
- Understanding the specific challenges of an Arabic NLP project
- Potentially can be expanded to summarizing Arabic articles

Data Source

We used 31,030 Arabic news articles and headline pairs in the Saudi Newspapers Arabic Corpus, SaudiNewsNet, that were extracted from 14 different news websites covering topics like politics, local news, culture, and sports with the following fields:

field	description
source	the newspaper from which the article was extracted
url	a full URL to the article
date_extracted	a timestamp of the date on which the article was extracted
title	the headline of the article (can be empty)
author	the author of the article (can be empty)
content	the body of the article

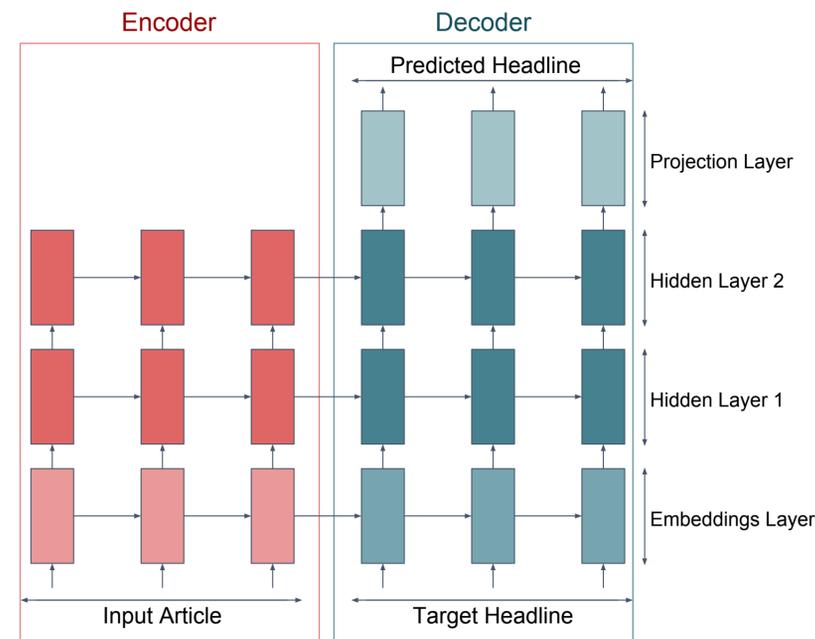
Data Preprocessing

- Data cleanup: remove annotations and punctuation marks
طالب → طالب
- Data splitting into 80% train, 8% validation and 12 % test sets
- Adding special tokens:
 - PAD: to pad articles and headlines below max length
 - UNK: for words not in the vocabulary
 - START and STOP: to start and stop the model decoder

Challenges

- The same Arabic word might be spelled differently due to annotation of conjugation of the word
- Training the model requires high computation power
- The complexity of the model made it harder to debug

Sequence-to-sequence model

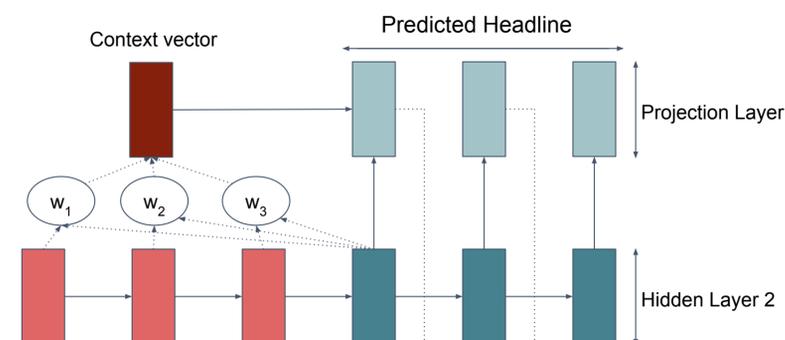


Hyperparameters

Learning Rate	Max Gradient Norm	Batch Size	Embeddings Size	Vocab Size	Article Length	Headline Length	Dropout Rate
1.0	5.0	64	64	100K	200	10	0.2

Adding Attention

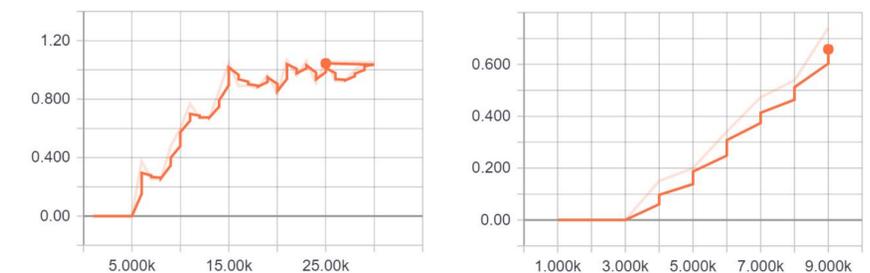
Motivation: Using attention can help when the input contains long sentences, such as in the case of articles.



Results

Results on numeric evaluation metrics:

Model	Dev		Test	
	Ppl.	BLEU	Ppl.	BLEU
Standard (30K)	3969	0.9697	3663	1.053
Standard (10K)	1570	0.6354	1512	0.6003
Attention	2700	0.6653	2417	0.7408



Standard (left) and Attention (right) models BLEU score on the test set vs. number of batches trained.

Examples and Observations

Examples:

Human title (translated):

6 Died and Injured in a Traffic Accident East of Quwaieyah

Standard title (translated):

8 of One Family Died and Injured in a Car Flip Over on Khurais

Attention title (translated):

Three Died and Seven Injured in a Traffic Accident to Arar

General Observations:

- Can't strongly affirm which model is better
- Generated titles overall seem related to the article
- They tend to make up some of the details mentioned in the article
- In the example given above, both of the models talk about the car accident. But, they also made up the numbers and the location of the accident. The standard model also decided all of these people are related and the accident was, in fact, a flip over

Next Steps

- Train the attention model for more epochs
- Try a copying mechanism to avoid unks and better incorporate article details (numbers, nouns, etc) in the title