



RED DRAGON AI

Unsupervised Natural Question Answering with a Small Model

Martin Andrews
martin@RedDragon.ai

Sam Witteveen
sam@RedDragon.ai



Summary

Task :

- ▲ Answer Natural Questions (factoids)

Builds upon :

- ▲ GPT2-117M model
- ▲ Unsupervised training throughout

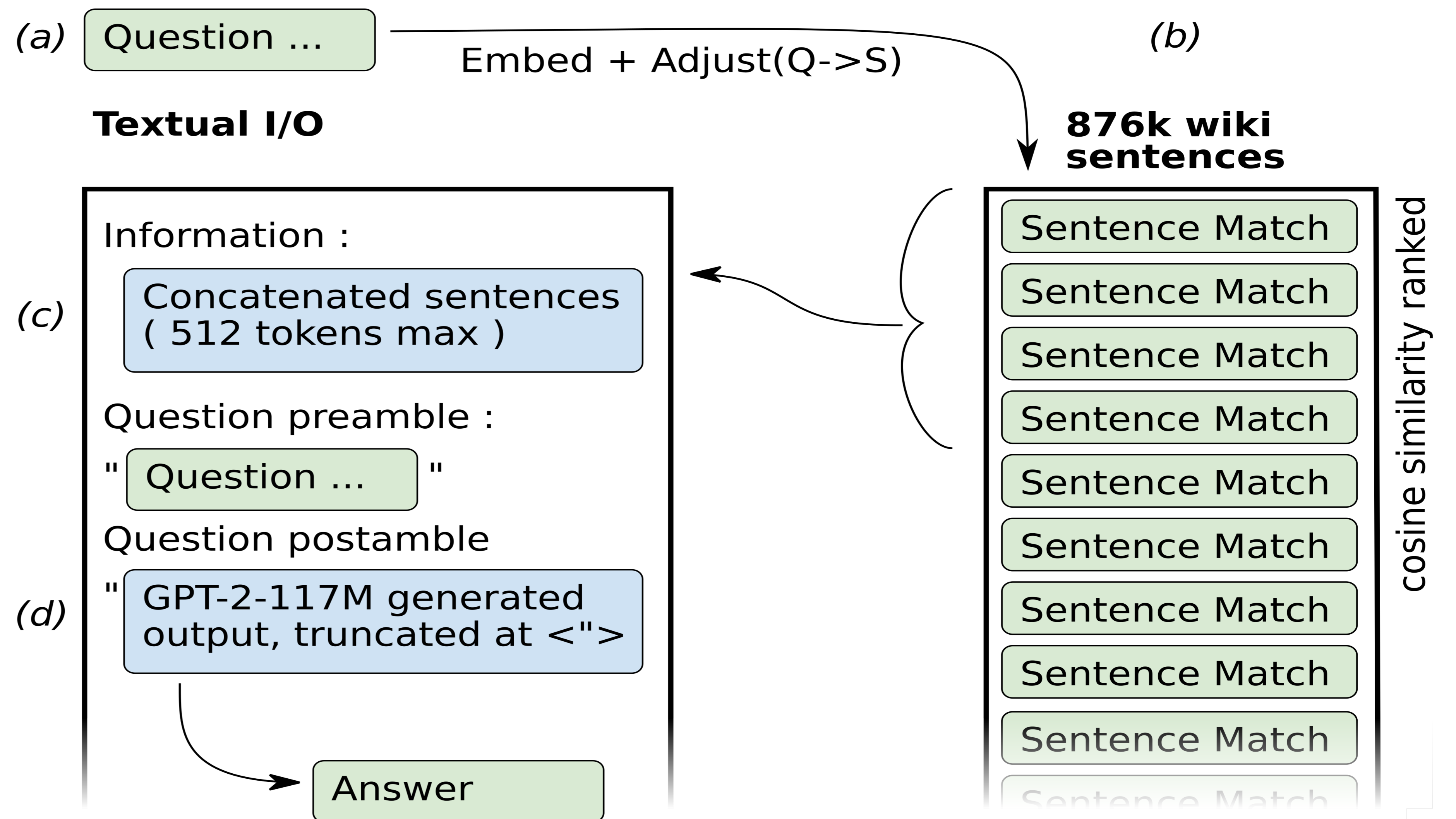
Ideas :

- ▲ GPT2-1.5Bn factoid results surprising
- ▲ Facts stored in weights ...
- ▲ Test 117M model combined with facts retrieved from Wikipedia

Results :

- ▲ Purely unsupervised approach
- ▲ Higher score than 1.5Bn model
- ▲ USE model (supervised) even better

Unsupervised Small Model QA : System Architecture



Dataset - Visual Genome (with MS COCO splits)

Sample question answers with filter examples, and examples of answers where pure SQuAD accuracy did not make sense when the base data included far more information than the original (single) wiki article targeted by the Natural Questions dataset.

Question	Target	GPT-2-117M	Reject reason
Who is the richest club in the championship?	'Aston Villa', 'Manchester City'	The richest club in the championship	SMART ALEC
Are all firestone tires made in the usa?	'NO'	No	Y/N QUESTION
What is the name of manchester united stadium?	'Old Trafford'	Manchester United	WITHIN QUESTION
Who cracked the enigma code in world war 2?	'Turing'	Alan Turing	N/A : ACCEPTED
How many inches is the iphone 5s screen?	'4 - inch screen size', '4 in', '4 in (10 cm)'	4 inches	N/A : ACCEPTED

Results

Question answering accuracy:

EMBEDDING	DIM	α	SCORE
NO HINTS	-	0.0	0.84%
BERT-REST	768	0.0	1.08%
SIF	768	0.7	3.14%
SIF	768	0.2	3.29%
USE	512	0.0	4.45%

- ▲ SIF = Smooth Inverse Frequency
- ▲ USE = Universal Sentence Encoder
- ▲ Comparable GPT2-1.5Bn Accuracy :
4.1% - (1.8% of Y/N) = 2.3%

Discussion

Basic Idea Validated :

- ▲ Unsupervised methods work
- ▲ Can use pure text as 'memory'
- ▲ Enabled far smaller LM to be used

Just a Proof-of-Concept :

- ▲ Storing facts in weights not necessary
- ▲ More efficient to query text itself
- ▲ Text processing is a text task

Future directions :

- ▲ Use text as 'internal signalling'
- ▲ Interpretability benefits
- ▲ Explore multi-stage processing
- ▲ Self-consistency checking

Source code available:

- ▲ <http://RedDragon.ai/research>

Key References

- "Language models are unsupervised multitask learners" - Radford et al. (2019)
- "Natural Questions : a benchmark for question answering research" - Kwiatkowski et al. (2019)
- "A simple but tough-to-beat baseline for sentence embeddings" - Arora et al. (2016)

Contact

martin@RedDragon.ai
+65 8585 1750
<http://RedDragon.ai>