GYM at Qur'an QA 2023 Shared Task: Multi-Task Transfer Learning for Quranic Passage Retrieval and Question Answering with Large Language Models

Ghazaleh Mahmoudi*, Yeganeh Morshedzadeh*, Sauleh Eetemadi

November 2023

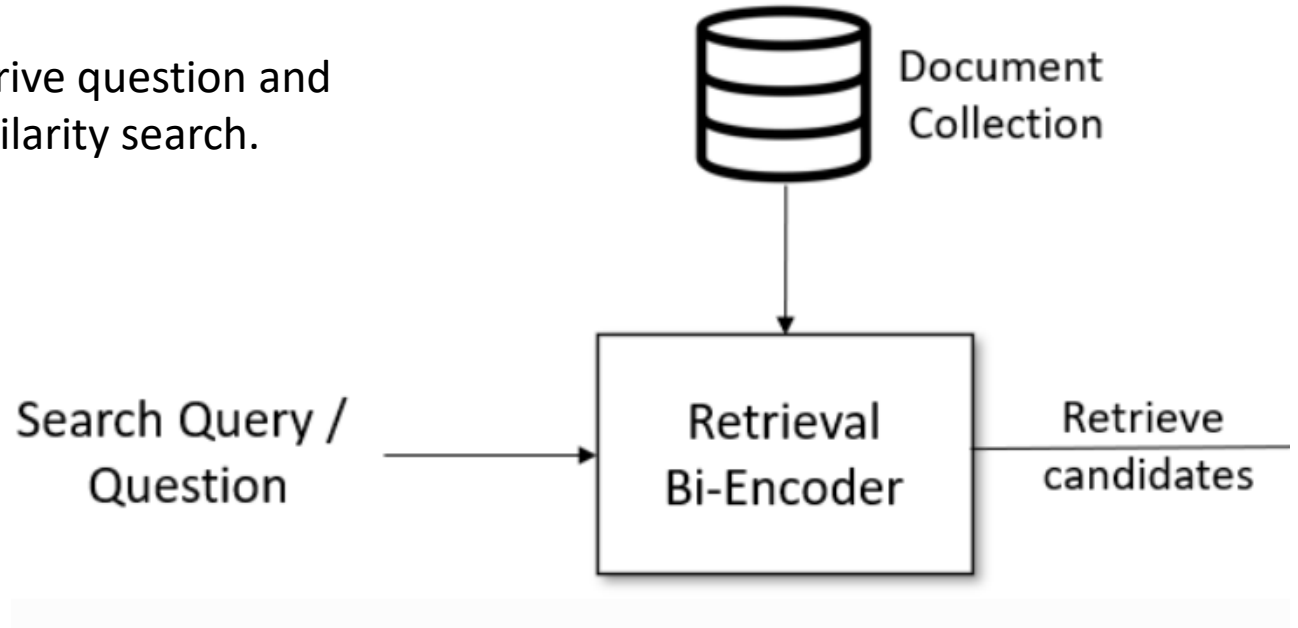
Overview

- Question-answering over the Quran
- Passage Retrieval
  - unsupervised fine-tuning of sentence embedding
  - supervised multi-task learning
- Reading comprehension
  - fine-tune an Electra-based model for QA

Task A : Passage Retrieval Approach

Unsupervised Fine-Tuning: Learning Sentence Embedding

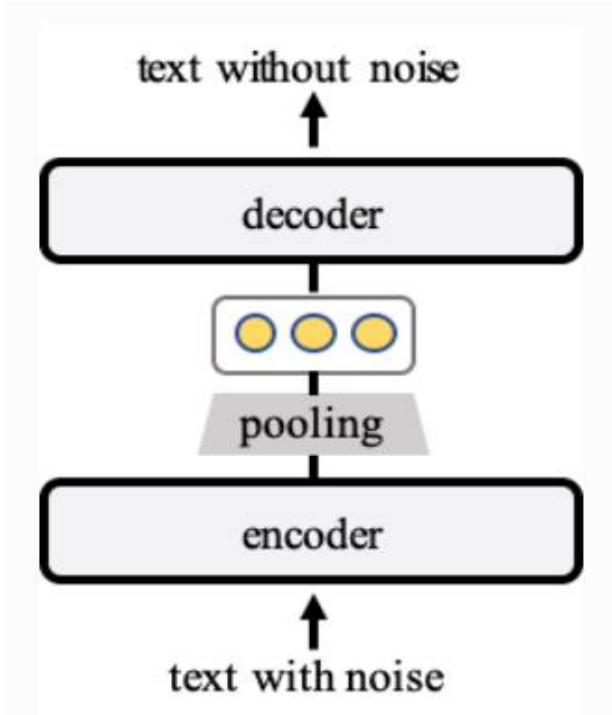
- Uses Sentence-BERT framework to derive question and passage embeddings for semantic similarity search.



Task A : Passage Retrieval Approach

Unsupervised Fine-Tuning: Learning Sentence Embedding

- Tranformer-based Denoising AutoEncoder (TSDAE)



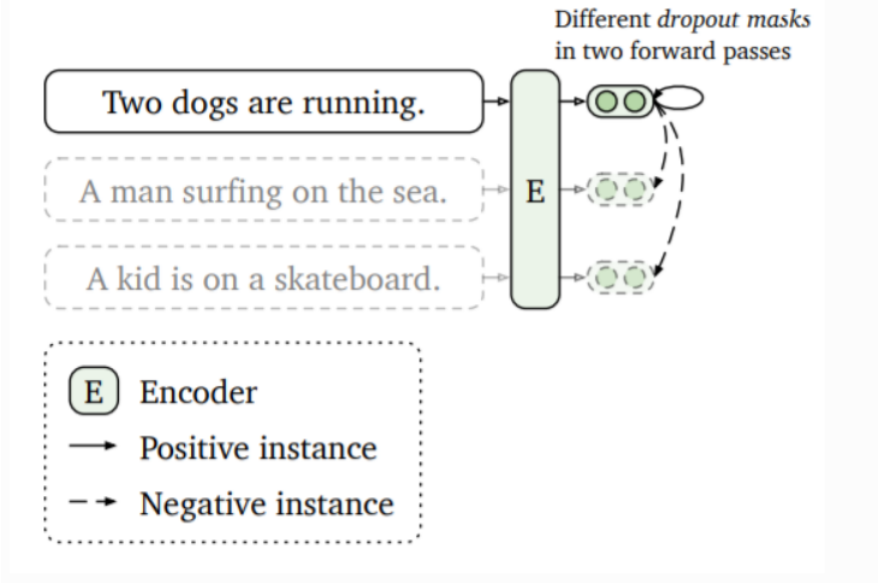
TSDAE

Kexin Wang, Nils Reimers, and Iryna Gurevych. 2021. TSDAE: Using transformer-based sequential denoising auto-encoderfor unsupervised sentence embedding learning. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 671–688, Punta Cana, Dominican Republic. Association for Computational Linguistics.
Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In Empirical Methods in Natural Language Processing (EMNLP).

Task A : Passage Retrieval Approach

Unsupervised Fine-Tuning: Learning Sentence Embedding

- Simple Contrastive Learning of Sentence Embeddings (SimCSE)



SimCSE

Kexin Wang, Nils Reimers, and Iryna Gurevych. 2021. TSDAE: Using transformer-based sequential denoising auto-encoderfor unsupervised sentence embedding learning. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 671–688, Punta Cana, Dominican Republic. Association for Computational Linguistics.
Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In Empirical Methods in Natural Language Processing (EMNLP).

Task A : Passage Retrieval Approach

Unsupervised Fine-Tuning: Learning Sentence Embedding

- Unsupervised: fine-tunes AraBERT with contrastive learning methods
  - SimCSE
  - TSDAE

Supervised Fine-Tuning: Training Bi-Encoder using Question-Passage Pairs

- Supervised: using multi-task learning to train a bi-encoder
- Mr. TyDi
  - Negative ranking loss:
    - <*question*, *positive passage*, *negative passage*>
- Quranic question-passage pairs
  - Using BM25 as negative passages
    - Contrastive loss:

      <*question*, *passage*, *label*>
    - Triplet loss:

      <*question*, *positive passage*, *negative passage*>

Task A : Passage Retrieval Approach

		AraBERT-TSDAE-Contrastive	AraBERT-SimCSE-Contrastive	AraBERT-SimCSE-Triplet
Sentence Embedding	TSDAE	✓		
	SimCSE		✓	✓
Training Loss	Denoising Auto-Encoders	✓		
	Contrastive	✓	✓	
	Triplet			✓
	Multiple Negative	✓	✓	✓
Dataset	Quran Question-Passage	✓	✓	✓
	Mr TyDi	✓	✓	✓

Task A Models Summery

Task A : Passage Retrieval Approach

Model Name	Train Set		Development Set		Test Set	
	MAP	MRR	MAP	MRR	MAP	MRR
AraBERT-TSDAE-Contrastive	0.1502	0.3206	0.1365	**0.2613**	**0.0545**	**0.1581**
AraBERT-SimCSE-Contrastive	**0.6522**	**0.7646**	**0.1459**	0.2573	0.0315	0.1023
AraBERT-SimCSE-Triplet	0.5243	0.6580	0.1082	0.1693	0.0116	0.0356

Task A MAP@10 and MRR@10 Results

Task B: Reading Comprehension

- Fine-tunes AraElectra models pre-trained on Arabic using SQuAD and TyDiQA datasets

	Dataset			Model and Environment Setting			
	SQuADv2	TyDiQA	QRCD v1.2	Epoch	Batch Size	Max sequence Length	Document Stride
AraElectra-SQuADv2	✓		✓	30	4	256	64
AraElectra-TyDiQA		✓	✓	1	8	256	64

Task B train setting

Task B: Reading Comprehension

- AraElectra-SQuADv2
- AraElectra-TyDiQA
- Ensemble Modeling
  - Same answer sum the model's output scores
  - Different Answer keep answer without changing the score.
  - calculated scores, sort the output, select the top 10 outputs as final result

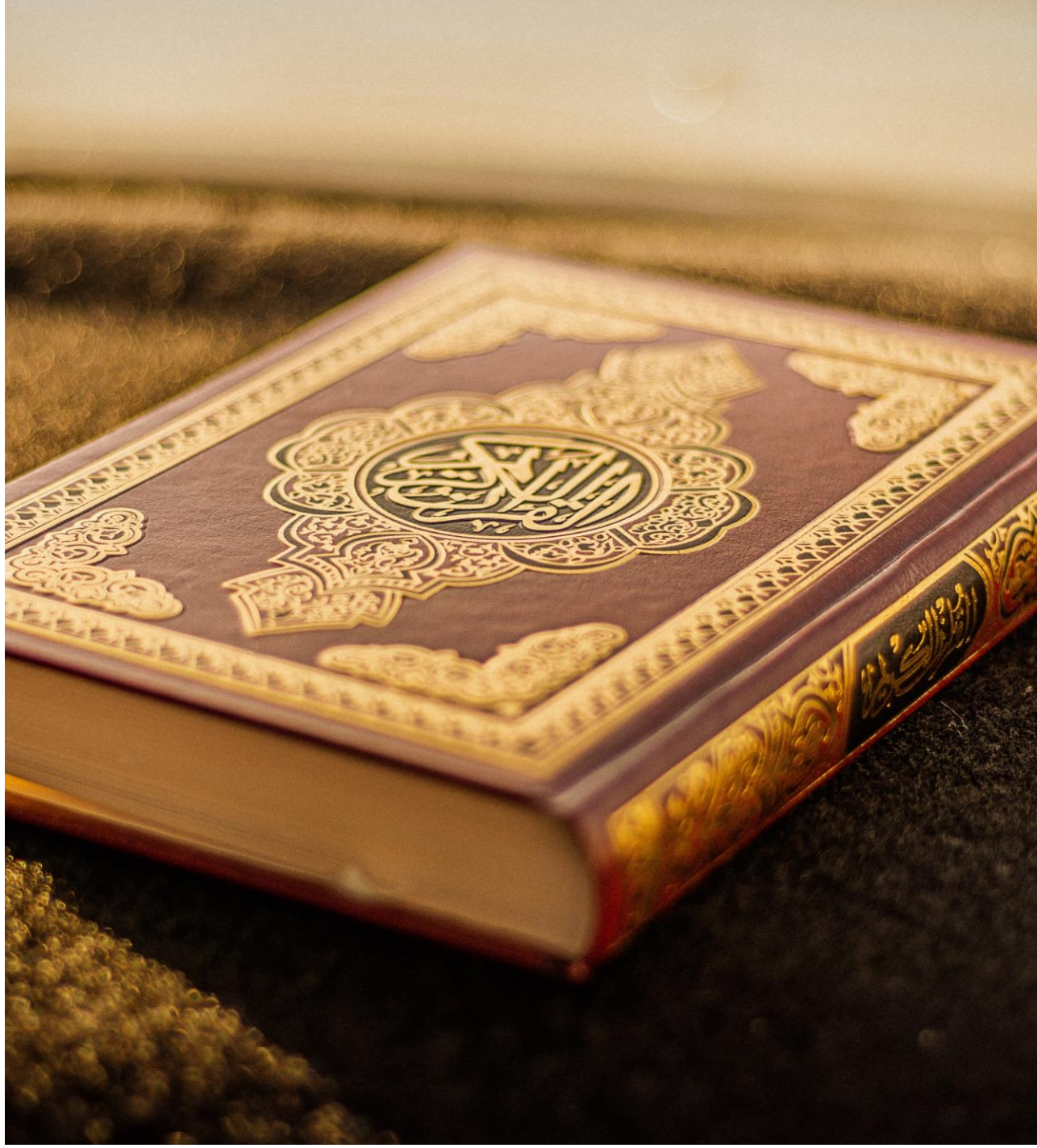
Task B: Reading Comprehension

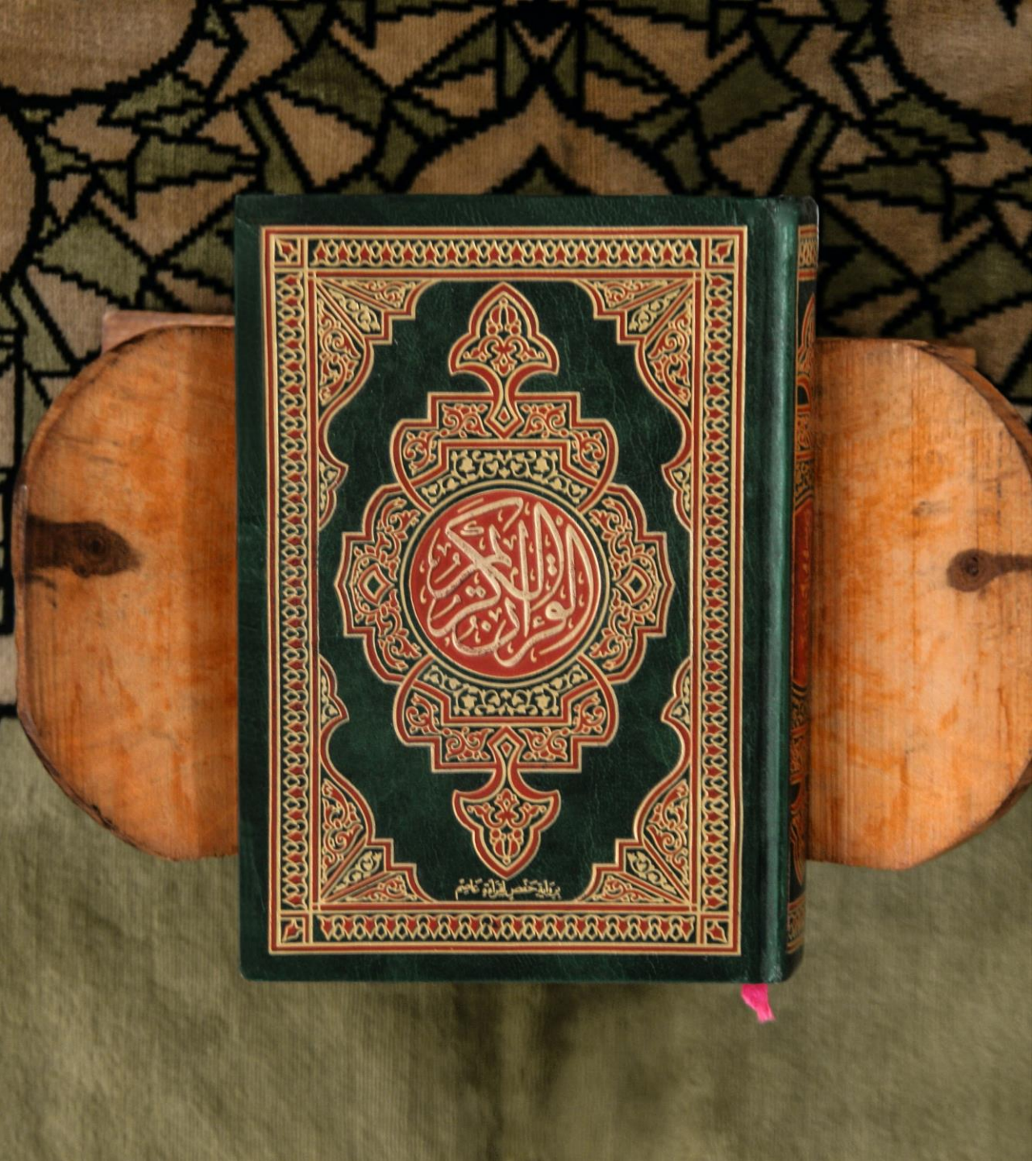
Model Name	Development Set	Test Set
AraElectra-SQuADv2	**0.485**	**0.461**
Ensemble	0.481	0.458
AraElectra-TyDiQA	0.431	0.430
Baseline	0.255	0.326

Task B pAP@10 result

Conclusion

- Transfer learning is effective despite limited Quranic training data
- Pre-training provides useful linguistic knowledge, while fine-tuning specializes models
- Techniques like multi-task learning further improve performance





Thank you

gh_mahmoodi@comp.iust.ac.ir

yeganeh.morshedzadeh@ubc.ca

sauleh@iust.ac.ir