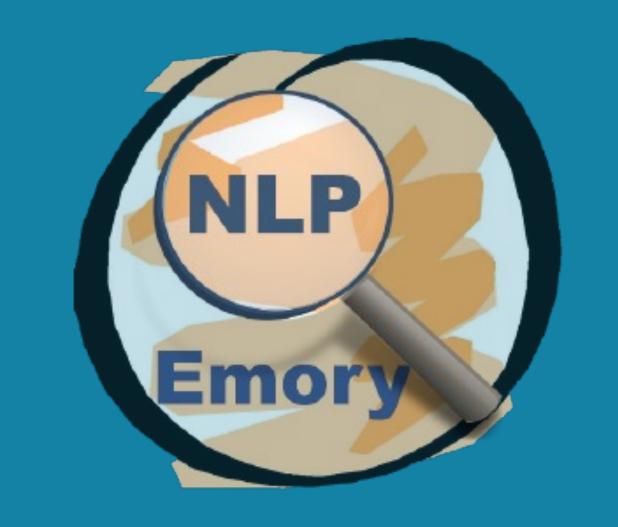


When Large Language Models Meet Religious Text

Jacob Choi, Jinho Choi Emory University Department of Computer Science



Introduction

Motivation:

Our goal is to explore methods to create open-source tools that can help empower religious communities by answering questions they have from a computational perspective.

Commercial Religious Al Tools:







Objectives:

Employ semantic similarity search to find reference verses in the Bible. Use Maximum weight matching algorithm to tackle limitations from semantic similarity search in finding references.

Using a state-of-the-art large language model to train on 6 datasets we made to answer user questions.

Semantic Similarity Search

Verse (John 12:13) The crowds that went ahead of him and those that followed shouted "Hosanna to the Son of David!" "Blessed is he who comes in the name of the Lord!" "Hosanna in the highest heaven!"

1.)Build Embedding 2. Measure Similarity 3. Pairwise Matching 4. Match top-k

Verse (Psalm 118:26) Blessed is he who comes in the ame of the LORD. From the house of the LORD we bless you.

Sentence Transformer all-MiniLM-L6-v2

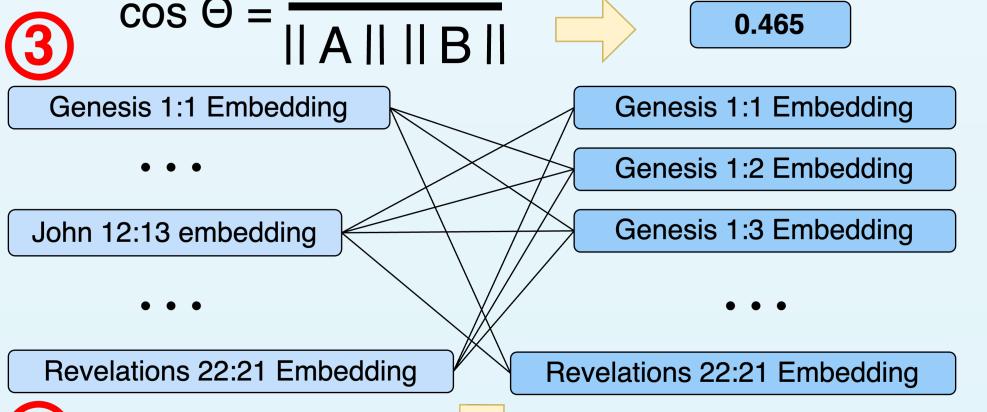
Embedding (John 12:13): [5.0864e-03, 1.9370e-01,

Embedding (Psalm 118:26): [1.5638e-02, 1.0540e-01, 4.4104e-02 ...]



-4.9242e-02, ...]





John 12:13 Mark 11:9 (0.8089)

Matthew 21:9 (0.8006)

• • •

Psalm 118:26 (0.465)

Bible Entities 0.845 Mark 15:18 (0.7464) **Matthew 3:6**

Confessing their sins, they were baptized by hin in the Jordan River.

 \mathbb{P}^{\square}

Entities: "the Jordan River", "LOC"

Create questions a user may have using entities:

Maximum Weight Matching

Psalm 118:26

Cosine Similarity

+ Squaring

3

Verse (John 12:13) Verse (Psalm 118:26) The crowds that went ahead of Blessed is he who him and those that followed comes in the name of shouted, "Hosanna to the Son of the LORD. From the David!" "Blessed is he who house of the LORD we comes in the name of the bless you. Lord!" "Hosanna in the highest heaven!"

John 12:13

Extracted Key Phrases:

in the name of the LORD. From the house of the LORD

Blessed is he who comes

we bless you.

Sentence Transformer all-MiniLM-L6-v2



Key Phrase Embeddings:

• [0.0212, 0.1114, -0.0506, ...] • [0.0096, 0.1038, 0.0230, ...] • [0.0218, 0.1564, 0.0202, ...]

• [0.0155, 0.0883, -0.0338, ...]

1. Extract Key

Squaring

2. Build Embedding

Measure Similarity

3. Pairwise Matching

.)Maximum weight

Extracted Key Phrases:

• Blessed is the king of Israel!

Shouting 'Hosanna!'

of the Lord!

Palm branches were taken to meet him.

Key Phrase Embeddings:

• [.01690, 0.1132, 0.0217, ...] • [0.0040, 0.0461, 0.0244, ...

Embeddings cosine similarities 0.0404 0.0361 0.0338 0.0568 John 0.9386 0.3028 0.4539 0.2316 cosine similarities Psalm 0.0404 0.0361 Square each cosine 0.0338 0.0568 John similarity

Total Samples

126,030

Total Samples

75,776

Total Samples

82,847

Total Samples

1,253

Total Samples

155,458

Total Samples

73,301

{"instruction": "What did God create in the beginning?", "output": "Genesis 1:1"}

{"instruction": "What will flow out of the LORD's house?", "output": "Joel 3:18"}

0.9386 0.3028 0.4539 0.2316 cosine similarities reference verses

Training

100,824

Training

60,620

Training

66,277

Training

1,002

Training

124,366

Training

58,640

{"instruction": "These things I command you, so that you will love one another.", "output": "John 15:17 ESV"}

{"instruction": "What is the cross reference for Luke 10:23?", "output": "Matthew 13:16-17"}

{"instruction": "What is the verse referenced in Matthew 5:21?", "output": "Exodus 20:13"}

{"instruction": "What does the verse in Mark 1:3 point to?", "output": "Isaiah 40:3"}

{"instruction": "This is my command: Love each other.", "output": "John 15:17 NIV"}

{"instruction": "What is the cross reference for Genesis 1:7?", "output": "Psalm 148:4, Proverbs 8:28-29"}

{"instruction": "What does the Bible say about God's omniscience?", "output": "1 John 3:20, Psalm 147:5"}

{"instruction": "How does the Bible instruct believers to dress modestly?", "output": "Deuteronomy 22:5"}

Evaluation

12,603

Evaluation

7,577

Evaluation

8,284

Evaluation

125

Evaluation

15,545

Evaluation

7,330

Matching John 12:13 0.0404 0.0361 0.0404 0.0361 0.0338 0.0568 0.0338 0.0568 John **0.9386** 0.3028 0.9386 0.3028 0.4539 **0.2316** 0.4539 0.2316 **Embeddings** cosine similarities Apply the maximum weight matching algorithm to obtain the numbers in red. Sum the red values to obtain the score. Matthew 21:9 (1.597) John 12:13 Mark 11:9 (1.305) Sim Max Psalm 118:26 (1.170) 184 202 • • • 223 231 SCORES: 235 243 241 251 Number of correct reference verses caught for semantic 25 250 261 similarity (Sim) compared to maximum weight matching 255 265 (Max) for different k out of 345

Test

12,603

Test

7,579

Test

8,284

Test

126

Test

15,545

Test

7,330

260

Avg Token Len

31.75

Avg Token Len

82.00

Avg Token Len

80.97

Avg Token Len

38.64

Avg Token Len

54.12

Avg Token Len

31.62

35

Maximum Weight

Datasets

Dataset Name

Single

Dataset Name

Situation

Dataset Name

Similarity

Dataset Name

References

Dataset Name

Version

value.

Bible verses scraped from BibleGateway.com.

Created 6 instruction-tuning datasets.

F#1

6,653 unique

topics

verses with

cross references

NIV: 314

NLT 167

Union: 345

Version:

Micro F1 on

Dataset Instructions created by prompting GPT 3.5-turbo-0125.

Single:

Come up with grammatically simple, factual questions whose answers can be found in: [verse]

Situation: Topical Bible utilized from OpenBible.info to extract topics. 30,536 samples

Prompt 2: Come up with life situation samples in 45,239 samples simple language that one may have from: [topic] in first person that can be answered from the Bible.

Similarity: Cross references extracted from OpenBible.info. 29,364 unique

Prompt: Come up with ways of asking what is the cross reference of [verse]

Come up with questions one may have about:

[topic] that can be found in the Bible.

References: Reference verses extracted from NIV and NLT footnotes.

[verse] is referencing a different verse. Create questions about how one might ask for the referenced verse mentioned in [verse].

Utilized NIV, NLT, KJV, NKJV, ESV versions. **NER**:

Entities extracted with Emory Language and Information Toolkit. **ELIT Macro &**

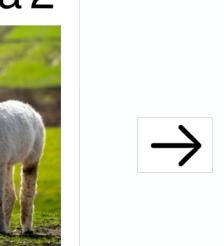
Dataset Name NER

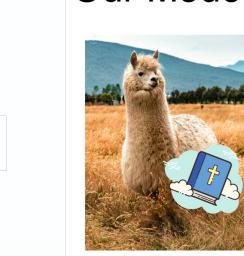
{"instruction": "Who was 105 years old when he became the father of Enosh?", "output": "Genesis 5:6"} {"instruction": "Who are the sons of Levi?", "output": "1 Chronicles 6:1"}

Datasets Our Model Llama 2

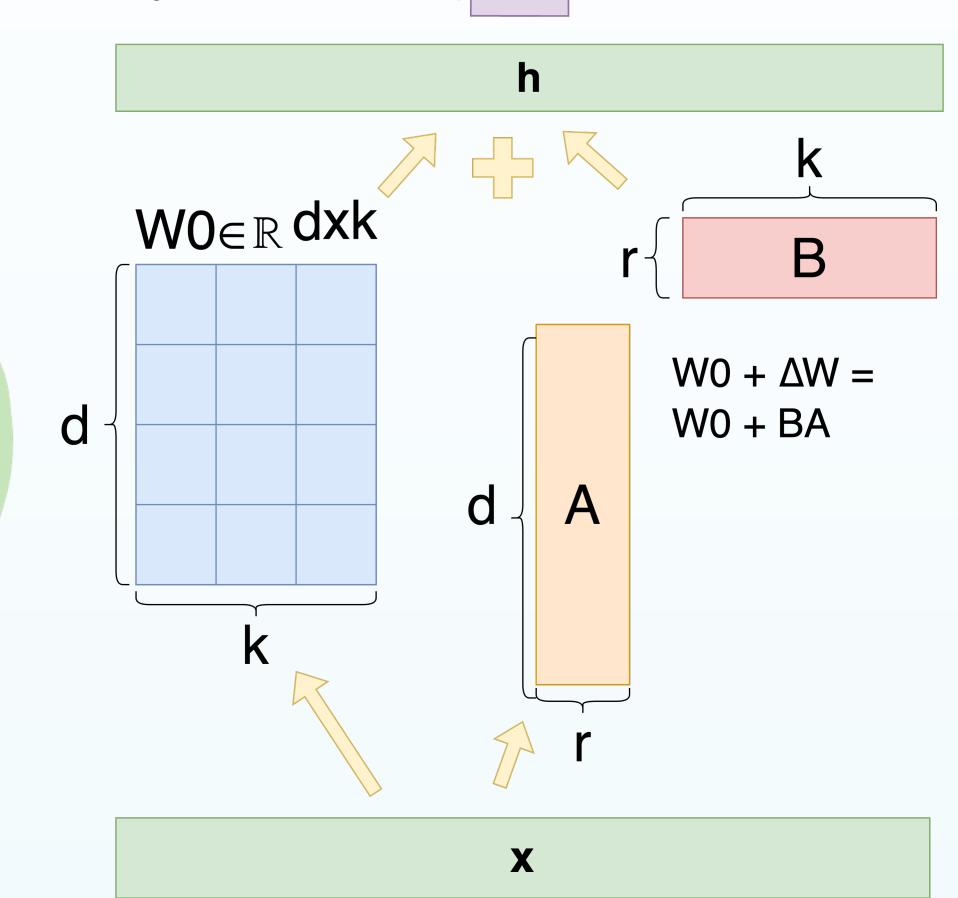
Fine-Tuning







Fine-tune a llama-2-7b-chat model on datasets using instruction fine-tuning and Low Rank Adaption.



Low rank adaption allows for memory efficient fine tuning on a single GPU

Results

NER	Situation	References	Single	Version	Similarity	Combined
.1752	0.3421	0.872	0.2187	0.4888	0.4961	0.3071

Exact Match Evaluation As Percentage

Exact match used as evaluation during training. Represents best exact match during iterations

Model	Christian Questions	Factual Questions
GPT 3.5-Turbo	24	15
Llama-2 Finetuned	5	1
Llama-2 Pretrained	18	9

Comparison of Performance on Christian and Factual Questions Dataset. Number correct out of 41 for Christian Questions, and number correct out of 24 for Factual Questions

Key Findings

- Discover novel approach using maximum weight matching to finding reference verses in the Bible that improves upon a sematic similarity search
- Commercial large language models are still superior to verse extraction task, though our models are transparent about their dataset task.