Enhancing Arabic Question Answering (QA) Systems Using Large Language Models (LLMs)

Supervisor : Mahran Farhat

Overview

Arabic Question Answering (QA) systems aim to provide precise and contextually accurate answers to questions posed in Arabic, spanning Modern Standard Arabic (MSA) and regional dialects. Despite advances in LLMs like GPT, mBERT, and BLOOM, Arabic QA systems face challenges due to the linguistic complexity, morphological richness, and dialectical diversity of the Arabic language. This research focuses on leveraging LLMs to improve Arabic QA systems by fine-tuning models, creating benchmarks, and exploring advanced techniques like few-shot and zero-shot learning.

Objectives

1. Fine-Tune LLMs for Arabic QA:

  - Adapt multilingual LLMs (e.g., XLM-R, mT5, GPT models) for Arabic-specific QA tasks.
  - Train models to handle diverse Arabic input, including dialects and MSA, across domains like healthcare, education, and news.

2. Develop Benchmarks for Arabic QA:

  - Create comprehensive Arabic QA datasets in various domains.
  - Establish standardized evaluation metrics for Arabic QA systems to measure accuracy and contextual understanding.

3. Explore Few-Shot and Zero-Shot QA:

  - Investigate how LLMs can answer Arabic questions with minimal task-specific training data (few-shot learning).
  - Utilize transfer learning and multilingual capabilities to address Arabic QA without explicit Arabic training data (zero-shot learning).

Key Challenges

- Dialect Diversity: Addressing the variability in vocabulary, syntax, and semantics across Arabic dialects.
- Data Scarcity: Lack of large-scale, annotated datasets for Arabic QA tasks.
- Morphological Complexity: Managing Arabic's rich morphology and contextual ambiguity.

Research Plan

1. Dataset Preparation:

   - Utilize existing Arabic QA datasets such as TyDi QA, MLQA, or ARCD (Arabic Reading Comprehension Dataset).
   - Augment datasets by extracting QA pairs from Arabic text sources like Wikipedia, news articles, and social media.

2. Model Fine-Tuning:

   - Fine-tune pre-trained models like mBERT, AraBERT, or GPT for Arabic QA tasks.
   - Experiment with domain-specific fine-tuning for areas like healthcare or education.

3. Model Evaluation:

   - Develop benchmarks using standardized metrics (e.g., F1 score, exact match).
   - Evaluate performance across MSA and dialectical Arabic inputs.

4. Few-Shot and Zero-Shot Techniques:

   - Apply prompt engineering to leverage multilingual LLMs for Arabic QA without explicit fine-tuning.
   - Use adapters or lightweight models to enhance performance in few-shot settings.

Potential Applications

- Customer Support: Arabic-speaking virtual assistants and chatbots for businesses.
- Education: Arabic QA tools for e-learning platforms.
- Healthcare: Arabic medical QA systems for patient interactions.

References and Resources

1. Datasets:

   - Arabic Reading Comprehension Dataset (ARCD):
   https://github.com/husseinmozannar/ARCD
   - TyDi QA: https://github.com/google-research-datasets/tydiqa
   - MLQA: https://github.com/facebookresearch/MLQA

2. Pre-Trained Models:

   - AraBERT: https://github.com/aub-mind/arabert
   - XLM-R: https://huggingface.co/transformers/model_doc/xlmroberta.html
   - mBERT: https://github.com/google-research/bert

3. Papers:

- Mozannar, H., et al., "Neural Arabic Question Answering." Proceedings of ACL.

- Artetxe, M., et al., "Multilingual Question Answering Using Cross-Lingual Contextualized Representations." EMNLP.

- Abdelrahman et al., "ArabicaQA: A Comprehensive Dataset for Arabic Question Answering", Published in Annual International ACM... 26 March 2024 Computer Science, Linguistics

-Samah et al., "Arabic Question-Answering System Based on Deep Learning Models', In book: Innovations in Machine and Deep Learning (pp.133-159), 2023.

4. Tools:

- Hugging Face Transformers: For fine-tuning and deploying LLMs.
- OpenAI API: For exploring zero-shot and few-shot QA capabilities.