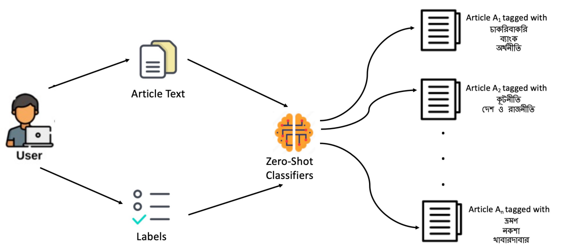
LLMs as Zero-Shot Multi-Label Classifiers for Bangla Documents

Souvika Sarkar, Shubhra Kanti Karmaker Santu
Big Data Intelligence (BDI) Lab, Auburn University

Motivation

- Bangla, the sixth most spoken language globally, poses NLP challenges due to its complexity and limited resources.
- Large Language Models (LLMs) like GPT, BLOOM, and LLaMA have transformed multilingual NLP tasks, their success primarily shines in widely spoken languages such as English and Chinese.
- The effectiveness of these models in low-resource languages like Bangla remains largely unexplored, motivating our study.
- Our Work explores the performance of state-of- the-art sentence encoders, and four LLMs in Zero-Shot multi-label classification (Zero-Shot-MLC) tasks.

Zero-Shot Multi-Label Classifiers



Benchmark Dataset

Dataset Name	# of Articles	Avg. article length	Labels retained	Labels/ article
BanglaNewsNet	7245	≈2517 words	21	1.345

Website:

https://www.prothomalo.com

- We created a new benchmark corpus, by crawling a large collection of publicly available online news.
- Each article here is already labeled with one or more labels by human annotators.
- A sample article: মেসিকে রিয়াল বেতিসে চান তাঁর আর্জেন্টিনা দলের সতীর্থ (Messi's Argentina team-mate wants him in Real Betis) is associated with labels "ফুটবল" (Football) and "আর্জেন্টিনা ফুটবল দল" (Argentina Football Team).

Sentence Encoders and LLM

Sentence Encoders : 1) Language-Agnostic SEntence Representations (LASER), 2) Language-agnostic BERT Sentence Embedding (LaBSE), and 3) Bangla sentence embedding transformer.
Large Language Model : 1. BLOOM, 2. FLAN-UL2, 3. GPTNeoX, 4. ChatGPT

Problem Statement

- **Definition 1.** 0SHOT-MLC: Given a collection of documents denoted as D = $\{d_1, d_2, ..., d_n\}$, a user represented by U and a set of user-defined labels denoted as $L_U = \{l_1, l_2, ..., l_m\}$ provided in real-time, classify each document $d_i \in D$ with zero or more labels from L_U , without further fine-tuning.
- Steps used in our 0-shot-MLC approach:
  1. Input Document
  2. Embedding Generation
     - Article Embedding: We embed the entire article with sentence encoders and LLMs in a single shot.
     - Label Embedding: Different approaches discussed in next section.
  3. Threshold-based Label Assignment
  4. Zero-Shot multi-label classification

Label Embedding Approaches

- **Label-Name-Only**
  - **Encode only the label name/phrase.**
- **Label + Keywords**
  - **Encode both label name and keywords, then average all embeddings to generate the final label embedding.**
- **Label + Keyword +Definition**
  - **Extract the label's and keyword's definitions, encode these definitions separately using sentence encoders, and then average all embeddings to generate the final label embedding.**
- **Explicit-Mentions**
  - **First, extract all the articles explicitly mentioning the label/phrase using algorithm for all labels. Then, for each label, generate embeddings of all articles that are explicitly annotated/tagged with that label, then average them to obtain the ultimate label embedding.**

Sample ChatGPT Prompt

Prompt Design	
System setup	
The AI assistant has been designed to understand and categorize user input by the given labels. When processing user input, the assistant must predict the labels from one of the following pre-defined options: 'চাকরিবাকরি' (Job market), 'করোনাভাইরাস' (Coronavirus), 'চলচ্চিত্র ও তারকা' (Movies and celebrities), 'স্বাস্থ্য' (Health), 'ব্যাংক' (Banking), 'অর্থনীতি' (Economy), 'শিক্ষা' (Education), 'প্রাকৃতিক দুর্যোগ' (Natural disasters), 'আইন ও আদালত' (Law and justice), 'কূটনীতি' (Politics), 'শিল্প ও বাণিজ্য' (Industry and commerce), 'ভ্রমণ' (Travel), 'নকশা' (Design), 'ফুটবল' (Football), 'খাবারদাবার' (Food and dining), 'দেশ ও রাজনীতি' (Country and politics), 'আন্তর্জাতিক' (International), 'দেশের খবর' (Country news), 'রাশিয়া ইউক্রেন সংঘাত' (Russia-Ukraine conflict), 'ক্রিকেট' (Cricket), 'নারী' (Women). It is essential to note that an article may have multiple labels. If the user input is not relevant with any labels, the assistant should print nothing, indicating that the input does not align with the available categories. The agent MUST response with the following json format: {**"Labels": ["List of labels"**]}	
User	**Taking into account the given Bangla article** {ব্যাংক এশিয়া ২০১৪ সালে ব্যাংকিং সেবার বাইরে থাকা বিপুল জনগোষ্ঠীকে ব্যাংকিং সেবায় আনতে এজেন্ট ব্যাংকিং সেবা চালু করে। বর্তমানে রাষ্ট্রমালিকানাধীন ও বেসরকারি মিলিয়ে ৩১টি ব্যাংক এ সেবা দিচ্ছে। বর্তমানে এজেন্ট ব্যাংকিং সেবা গ্রহণকারীর সংখ্যা দেড় কোটির বেশি। এর মধ্যে ব্যাংক এশিয়ার গ্রাহক ৫৫ লাখের বেশি। এসব গ্রাহকের ৯২ শতাংশই গ্রামীণ জনগোষ্ঠী। আবার ৬২ শতাংশ গ্রাহকই নারী। সারা দেশে ব্যাংক এশিয়ার ৫ হাজার ৪০০-এর বেশি এজেন্ট আউটলেট রয়েছে, যাদের মধ্যে নারী এজেন্ট ৫৪০ জন। ..... (In 2014, Bank Asia introduced agent banking services to bring banking services to a large population that was outside the reach of traditional banking. At present, a total of 31 banks, including both public and private, are providing these services. The number of users availing agent banking services has surpassed 140 million. Among them, Bank Asia has more than 5.5 million customers, of which 92% are from rural areas. Furthermore, 62% of the customers are women. Throughout the country, there are more than 5,400 agent outlets of Bank Asia, including 540 outlets managed by female agents.....)}, predict the category or labels of this article from the list of mentioned labels.
ChatGPT	{"**Labels**": "ব্যাংক" (Banking), "শিল্প ও বাণিজ্য" (Industry and commerce)}
Directive: Taking into account the given Bangla article {article text}, predict the category or labels of this article from the list of mentioned {labels}. Please remember to only respond in the predefined JSON format without any additional information.	

Experiment Results

Sentence Encoder

Topic+Keywords Based Label Embedding								
LASER			LaBSE			BanglaTransformer		
Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1
0.162	0.750	0.267	0.282	0.477	**0.354**	0.224	0.648	0.334
Explicit-Mention Based Label Embedding								
LASER			LaBSE			BanglaTransformer		
Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1
0.193	0.724	0.305	0.300	0.617	**0.404**	0.276	0.635	0.384

Large Language Model

Topic+Keywords Based Label Embedding								
FLAN-UL2			BLOOM			GPT-NeoX		
Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1
0.135	0.890	0.234	0.231	0.574	0.329	0.235	0.634	0.345
Explicit-Mention Based Label Embedding								
FLAN-UL2			BLOOM			GPT-NeoX		
Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1
0.144	0.742	0.241	0.232	0.642	0.341	0.241	0.675	0.357

ChatGPT

Precision	Recall	F_1 Score
0.515	0.573	0.537

Discussion

- This paper assesses the effective- ness of contemporary LLMs for the Zero-Shot- MLC task exclusively for a widely spoken yet low- resource language, i.e., Bangla.
- Among the large sentence encoders, ChatGPT performed the best, followed by PaLM. However, GPT-NeoX and BLOOM didn't generalize effectively for the task.
- Our research contributes to the ongoing efforts to enhance the applicability and efficacy of LLMs for regional and low-resource languages, paving the way for future advancements in multilingual NLP research.