Effect of Pivot Language and Segment-Based Few-Shot Prompting for Cross-Domain Multi-Intent Identification in Low Resource Languages

Kathakali Mitra , Aditha Venkata Santosh , Soumya Teotia , Aruna Malapati
Department of CSIS
Birla Institue of Technology,Pilani,Hyderabad Campus

Abstract

NLU (Natural Language Understanding) has considerable difficulties in identifying multiple intentions across different domains in languages with limited resources. Our contributions involve utilizing pivot languages with similar semantics for NLU tasks, creating a vector database for efficient retrieval and indexing of language embeddings in high-resource languages for Retrieval Augmented Generation (RAG) in low-resource languages, and thoroughly investigating the effect of segment-based strategies on complex user utterances across multiple domains and intents in the development of a Chain of Thought Prompting (COT) combined with Retrieval Augmented Generation. The study investigated recursive approaches to identify the most effective zero-shot instances for segment-based prompting. A comparison analysis was conducted to compare the effectiveness of sentence-based prompting vs segment-based prompting across different domains and multiple intents. This research offers a promising avenue to address the formidable challenges of NLU in low-resource languages, with potential applications in conversational agents and dialogue systems and a broader impact on linguistic understanding and inclusivity.

Keywords : Retrieval Augmented Generation (RAG) , Chain of Thought Prompting (COT) , Ada Embedding , GPT 4 , Chroma Vector Database , Embedding , High Resource Language (HRL) , Low Resource Language (LRL) , Large Language Models (LLMs) , Pre-Trained Language Models (PLMs)

1 Introduction

Cross-domain Multi-Intent Classification is a crucial task in Natural Language Understanding with extensive applications in conversational agents and dialogue systems.For low resource language, this presents an intricate and pressing challenge.We conducted our research specifically on Indian languages, with Bengali being identified as a language with limited resources.This paper unveils a comprehensive methodology that leverages high-resource languages as a pivot language.The foundation of our approach relies on discovering high-resource languages that closely resemble the semantics of their low-resource equivalents.This study utilizes the notion of Retrieval-augmented Generation on Language Models (LLMs) to examine the impact of using a pivot language that is semantically comparable and selecting a small number of instances from high-resource languages to build context.The significance of storing the embeddings in the vector database has also been investigated.Further experiments in the research showed how the effect of chunking of cross-domain, multi-intent user-utterances affects the results from LLMs.Various experiments conducted as a part of the research offer a key to unlocking the complex semantics of low-resource languages. In essence, our research journey optimizes intent classification and domain categorization in low-resource languages, rendering it more inclusive, efficient, and adaptable to linguistic nuances and unexplored endangered low-resource languages.

Overall , we make the following contributions:

(1) Usage of a Semantically Similar Pivot HRL to perform cross-domain multiple intent identification for a LRL.This is achieved by leveraging the HRL embeddings and the metadata stored in a vector database for faster retrieval and indexing,hence facilitating Retrieval Augmented Generation in the LRL.

(2) Explore the impact of segment creation and demonstrate how varying values of k for each segment influence the prompts for a complicated user utterance that spans several intents across distinct domains.

2 Related Work

In the low-resource language context, cross-lingual multi-intent and multi-domain extraction remains underexplored, notably for Indian languages like Bengali. Our research aims to bridge these gaps by employing a chain of thought prompting and retrieval-augmented generation utilizing a vector database. We use a few shot examples from a corresponding HRL to generate a context and use it as a prompt. We have also explored the effect of the choice of few shot examples for each chunk from a user utterance. For low-resource languages, prompting has been proven to be a promising technique and has been shown to give better results. In (Halike et al., 2023), the authors introduced RelationPrompt, a zero-shot relation extraction approach which utilizes prompts to generate synthetic relation examples, which are then used to train a relation extractor for low-resource languages. (Huang et al., 2023) further enhances multilingual capabilities by appointing cross-lingual-thought prompting (XLT). However, the substantial requirement of computational resources led the way to further research. For intent classification in data-scarce environments, (Parikh et al., 2023) explores zero-shot and few-shot techniques which utilize external knowledge, such as word embeddings or taxonomies, to transfer knowledge between similar tasks. (Lu et al., 2022) further investigates the impact of prompt order on the performance of few-shot learning in large language models. However, their focus on English does not cater to the complexity of languages like Bengali. These lead the way to retrieval-augmented generation and vector databases to address these challenges. Our approach combines insights from (Liu et al., 2023) and (Wang et al., 2022) to predict multi-intent and multi-domain with high-resource references and enhance performance in low-resource languages through data augmentation. (Nie et al., 2023) outlines an approach to augment the context of prompts with semantically similar sentences retrieved from a high-resource language that enhances NLU models for low-resource languages, offering insights into effective retrieval-augmented prompts. (Ghosh and Caliskan, 2023) investigates the gender bias of ChatGPT across six low-resource languages, including Bengali. This underlines the importance of mitigating biases in AI systems. In (Guerreiro et al., 2023), the authors concluded that hallucinations are a significant challenge for large multilingual translation models and recommended that developers of multilingual translation systems should be aware of the issue of hallucinations and take steps to mitigate it. (An, 2023) utilizes pre-trained language models (PLMs) to learn text representation and generation capabilities on a large-scale unsupervised Tibetan corpus. (Jin et al., 2022) emphasizes parameter-efficient prompt learning, aligning with our intent and domain prediction goals. However, they may not be tailored to the cross-lingual and multi-intent prediction that we aim to incorporate. (Stylianou et al., 2023) further introduces domain-aligned data augmentation, crucial for accurate cross-lingual intent and domain prediction in low-resource languages. It may not focus on the chain of thought prompting and vector databases, which we consider integral to our approach.

3 Dataset

For retrieving semantically similar documents, Wikipedia articles pertaining to the freedom movement of Indian have been extracted for Indian genre languages - Hindi, Bengali, Marathi, Telugu, and Punjabi. For Retrieval Augmented Generation, a MASSIVE dataset is used. MASSIVE is a parallel dataset of > 1M utterances across 51 languages with annotations for the Natural Language Understanding tasks of intent prediction and slot annotation. Utterances span 60 intents and include 55 slot types. MASSIVE was created by localizing the SLURP dataset, composed of general Intelligent Voice Assistant single-shot interactions. For testing the proposed methodology, a test dataset comprising cross-domain and multiple intent user-utterance was used.

4 Methodology

This section focuses on the approach for cross-domain multi-intent identification for low-resource languages using high-resource languages as a pivot language.The approach is essentially divided into four primary components : (1) Identifying a High-Resource Language (HRL) that is semantically similar to a Low-Resource Language (LRL) within a language family. (2) Implementation of a Vector Database to store embeddings for the pivot language. (3) Developing a systematic approach to determine the appropriate number of documents per chunk for Retrieval Augmented Generation in the context of intricate user-utterances that span

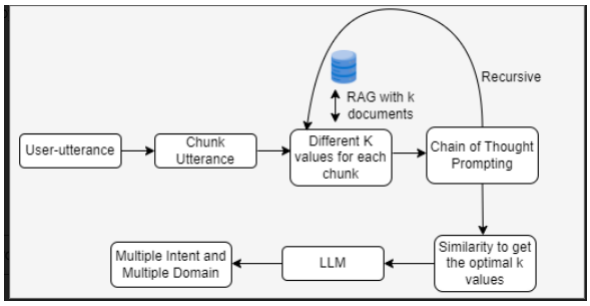

Figure 1: Proposed Architecture

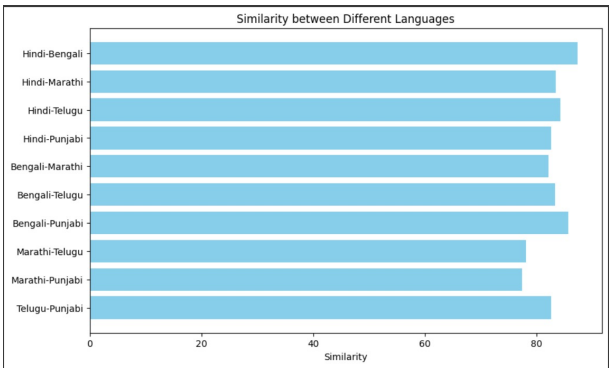

Figure 2: Semantic similarity in a language family. This diagram show the semantic similarity between HRL and corresponding LRL

across many domains and intentions, requiring segmentation into smaller chunks. (4) Conducting a comparison analysis to determine the accuracy and efficiency of segment-based prompting against sentence-based prompting (where a complete sentence is provided as a prompt without breaking it into smaller parts). For carrying out the experiments, text-embedding-ada-002 is used for extracting embeddings, and GPT 3.5 Turbo/GPT 4 is used as the large language model. Cosine Similarity is used as a similarity measure for computing similarity between documents. LangChain is used as an orchestrator to manage LLMs. Open Source In-Memory ChromaDB is used as a vector database for storing embeddings. The architecture diagram of the proposed methodology is given in Figure 1.

4.1 Semantically Similar Pivot HRL for a LRL

The proposed work focuses on a strategy to find the semantically similar high/medium resource language for a target low resource language. Previous experiments proved that semantically similar HRL can be used for different NLP tasks in LRL. Our experiment involved six Indian languages, and after careful evaluation, we determined that Bengali exhibited the highest semantic similarity with Hindi. We selected four Wikipedia articles with a common theme, focusing on the Freedom Movement of India, across the six languages: Gujarati, Bengali, Telugu, Tamil, and Marathi and conducted initial text preprocessing and tokenization. OpenAI's text-ada-embedding-002 was used for extracting the embeddings. We quantified the semantic similarity between different texts using cosine similarity, and our findings indicated that Hindi and Bengali have a higher degree of similarity. Hence Hindi was chosen as a pivot high/medium resource language for the target low resource language, Bengali.

4.2 RAG using Vector Database

The proposed research suggests the implementation of a vector database to store High-Resource Language (HRL) embeddings. These embeddings facilitate faster retrieval, indexing, and augmentation in the context of Retrieval Augmented Generation (RAG) when combined with Low-Resource Language (LRL) for the purpose of Natural Language Understanding (NLU) tasks, particularly for identifying multiple intents and domains.RAG (Retrieval-Augmented Generation) is instrumental in providing semantically similar documents in response to user queries. The methodology involves obtaining embeddings for HRL (Hindi) user utterances from the MASSIVE dataset using the text-ada-embedding-002 model. These embeddings are then stored in the Vector DB, along with relevant metadata such as intent, slots, and domains. ChromaDb serves as the platform for this database. The chunk_size chosen is 3000, and the chunk_overlap allowed was 200. For a user query in the target low-resource language, embeddings are computed using the text-ada-embedding-002 model. A cosine similarity search is then conducted to identify the top k relevant documents from the vector database, along with their associated metadata. The Langchain orchestrator is utilized to perform RAG on the vector database, generating context for the Chain of Thought Prompting, which is employed in zero-shot low-resource language intent and domain identification. This approach aims to enhance NLU tasks by leveraging embeddings and vector databases, enabling efficient document retrieval and generation based on user queries in low-resource languages.

4.3 Effect of Segment and k Values for RAG and COT Prompting

The proposed approach aims to find optimal combinations of zero-shot examples for HRL user-utterances to form the context in RAG for predicting multiple cross-domains and multiple intents for an LRL. It also establishes the effect of chunking the user-utterance into segments in performing better retrieval and forming efficient prompts. Since the user-utterances span across diverse domains, performing RAG with different values of k might lead to creating biased contexts, retrieving the data for a certain domain and missing on the other. It has been seen in the experiment conducted in the research that the zero-shot examples, which are biased towards a single domain, do not generalize well across all domains, adversely affecting the prompts for LLMs.

Given S = $x(i), (y(i), z(i))$ where $x(i)$ is user utterance and $(y(i)$ and $z(i)$) represents a tuple label comprising Domain and Intent. S is split into segments $c(i)$ where $c(i) \subseteq x(i)$ and $\bigcup(c(i)) = x(i)$. For the values of k from 2^i where $i > 0$ and $i \leq (UpperBound)$. We should ensure that tokens in the prompt do not exceed the maximum input capacity for an LLM. We have initially taken the whole sentence as one chunk and carried out our experiments for different values of k. The prompt created through RAG was sent to the LLMs for capturing multiple domains and multiple intents. A similarity measure captured the accuracy percentage against the target label.

Algorithm 1 Cross-Domain Multi-Intent Identification using RAG and COT for a single sentence passed as a one chunk

Input : user_utterance S
Output : intent_labels = [], domain_labels = []

foreach *k in [4, 8, 12, 16, 32, ...]* **do**
  $docs \leftarrow fetch\_docs(S, k)$
  $intent\_labels \leftarrow fetch\_metadata\_intent(user\_utterance, k)$
  $domain\_labels \leftarrow \_metadata\_domain(S, k)$
  $COT\_prompt \leftarrow (docs, intent\_labels, domain\_labels)$
  $res\_intent\_labels, res\_domain\_labels \leftarrow$ GPT$(prompt)$
return $res\_intent\_labels, res\_domain\_labels$

Algorithm 2 Cross-Domain Multi-Intent Detection Using Permutations to find the optimal k combination for COT

Input : s_1, s_2, s_3: Input sentences to be processed
Output : Intent and domain label for each k, for each chunk

sentences $\leftarrow [s_1, s_2, s_3, s_4, \ldots]$ result $\leftarrow$ []
foreach $sentence$ ***in*** sentences **do**
  $chunk\_list \leftarrow$ ***chunking***$(sentence, delimiter)$
permutations$(idx, permutation_intents)$
if $idx == chunk\_list.length$ **then**
  $score \leftarrow$ ***similarity***$(permutation\_intents, actual\_intent)$
  result.append$(permutation_intents, score, k)$
  return
foreach s_i ***in*** sentences **do**
  $chunk\_list \leftarrow$ ***chunking***$(sentence, delimiter)$
  foreach k ***in*** *[8, 12, 16, 32, ...]* **do**
    $c_i \leftarrow chunk\_list[idx]$
    $intent \leftarrow$ ***retrieve_get_intent***(c_i, k) _intents.$\{intent, k\}$
    permutations$(idx + 1, intent)$
    $permutation\_intents$.pop_back
permutations(*0, []*)
return result

5 Results and Discussion

The implementation of the model utilized Langchain, with ChromaDB serving as the vector database. The libraries developed by OpenAI were heavily utilized, and the specific model selected for this experiment was OpenAI's GPT-3.5-turbo.

5.1 Identification of Semantically Similar LRL corresponding to a HRL

We selected Bengali as our preferred LRL. We conducted an experiment including all the prominent Indian languages to determine which language is the most semantically related to Bengali. The Wikipedia entries on Freedom movements were retrieved from the Wikipedia API in the languages of our choice, specifically Hindi, Marathi, Telugu, and Punjabi. Text Pre-Processing was conducted and we ensured that the sentences had a length of 2048 or less. Additionally, we eliminated any null strings or NoneType strings from the pre-processed text. The text corpus was processed using OpenAI's text-embedding-ada-002 model to generate word embeddings. The ultimate semantic similarity was derived by calculating the cosine similarity between the Bengali text corpus and other corpora. The results are shown in Figure 2.

From the cosine similarities, we safely concluded that the most semantically similar language to Bengali is Hindi.Due to the abundance of data accessible for Hindi, we have selected Hindi as the pivot language in our case study.

5.2 Comparing LLMs for the Experiment

We conducted an experiment involving Llama2-13b and GPT3.5 turbo. A test dataset was subjected to Simple Random Sampling, and thereafter, LLMs were employed for Multi-Intent categorization. A similarity metric was plotted against the LLMs between actual and expected output, the results of which show GPT 3.5 Turbo performs well for the Bengali and Hindi utterances for the given NLU task for multiple domains and multiple intent identification.

From the above results, it was concluded that GPT-3.5 turbo gave better similarity scores for our test dataset, hence other experiments have been carried out using the same LLM.

5.3 Creating a Vector Database

The data from the corresponding HRL was used to create Langchain documents. The chunk_size chosen is 3000 and the chunk_overlap allowed was 200. The user utterance from the dataset was chosen as the page_content and other column labels like intent, slot_labels, domain and annot_utt were taken as metadata. The document embeddings were generated using the 'text-embedding-ada-002' model from OpenAI, and the resulting vectors were saved in ChromaDB..

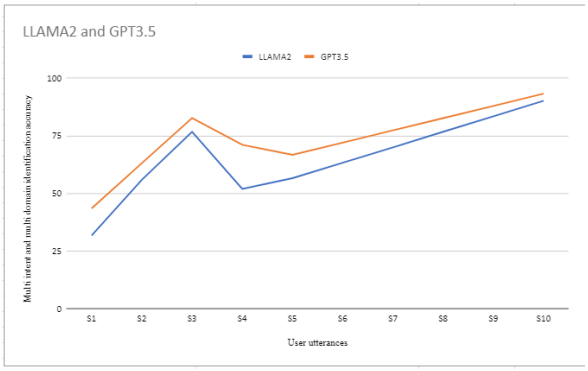

Figure 3: Comparison of LLM for multi-intent and multi-domain identification

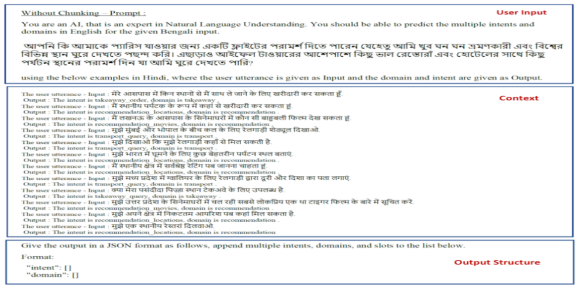

Figure 4: Chain of Thought prompting explained for k=8 semantically zero shot examples retrieved in HRL(Hindi) from vector database without chunking, where one sentence is one chunk

5.4 RAG and COT with Different Values of k Without Chunking

The experiment, as shown in Algorithm 1, was performed on the test dataset. A chain of thought prompt template is illustrated in Figure 4 for the given utterance in Bengali.

5.5 RAG and COT with Different Values of k with Chunking

User utterances from the test dataset were split into segments at each delimiter. For each segment, multi-domain and multi-intent classification was carried out based on the algorithm explained in Algorithm 2, for different values of k. The few shot examples for the optimal value of k for each chunk

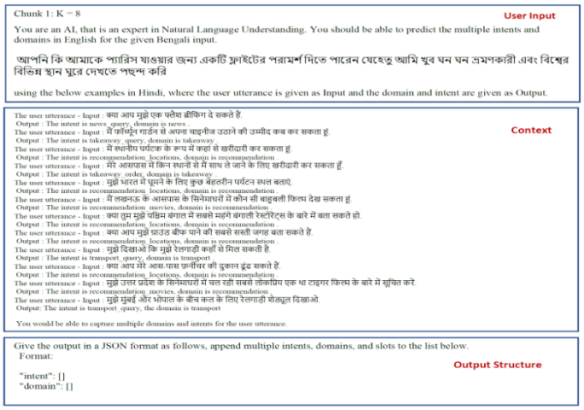

Figure 5: Chain of Thought prompting explained for k=8 semantically zero shot examples retrieved in HRL (Hindi) from vector database with chunking, where the sentence was split into two chunks. This diagram refers to the first Chunk of the user utterance in Figure 4

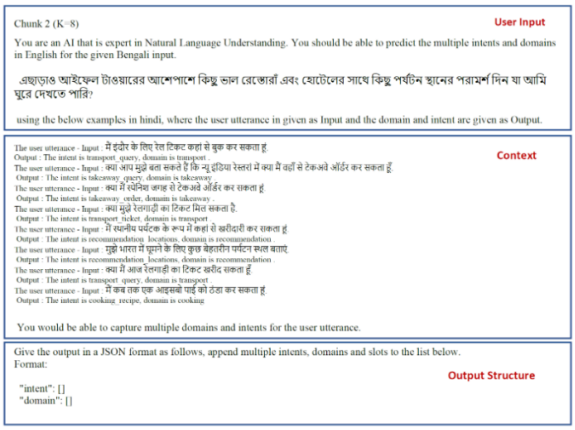

Figure 6: Chain of Thought prompting explained for k=8 semantically zero shot examples retrieved in HRL(Hindi) from vector database with chunking, where the sentence was split into two chunks. This diagram refers to the second Chunk of the user utterance in Figure 4

is given in figure 5 and 6.

5.6 Findings from LLM

We conducted experiments with various exponential values of k for each chunk and identified an optimal combination of k values by evaluating the similarity score of each combination with the expected intent and domain. Figure 7 provides an illustration of a complete sentence and its corresponding chunks. Figure 8 illustrates the results obtained with LLM GPT-3.5-turbo, particularly for the optimal values of k, corresponding to each segment of the sentence. Figure 9 demonstrates the outcomes obtained with LLM GPT-3.5-turbo for the optimal values of k, where each sentence is treated as a single unit.

A similarity matrix is plotted for each of the recursive combinations of chunks for k values. The top five similarities are shown below. Table. 1 shows the accuracy for a combination of an example input for each value of k. In the provided example, a total of 81 combinations were generated, and among them, this particular combination yielded the most favorable results.A comparison analysis was conducted, using RAG with COT, for one sentence treated as a single chunk with optimal values of k, and another sentence treated as several chunks for an ideal combination of k. The study clearly shows an improvement in accuracy of 4 to 5 % in the proposed segment-based prompting for NLU tasks of cross-domain multiple intents identification in low-resource language.

User Utterance : আপনি কি আমাকে প্যারিস যাওয়ার জন্য একটি ফ্লাইটের পরামর্শ দিতে পারেন যেহেতু আমি খুব ঘন ঘন ভ্রমণকারী এবং বিশ্বের বিভিন্ন স্থান ঘুরে দেখতে পছন্দ করি। এছাড়াও আইফেল টাওয়ারের আশেপাশে কিছু ভাল রেস্তোরাঁ এবং হোটেলের সাথে কিছু পর্যটন স্থানের পরামর্শ দিন যা আমি ঘুরে দেখতে পারি?

Chunk 1 : আপনি কি আমাকে প্যারিস যাওয়ার জন্য একটি ফ্লাইটের পরামর্শ দিতে পারেন যেহেতু আমি খুব ঘন ঘন ভ্রমণকারী এবং বিশ্বের বিভিন্ন স্থান ঘুরে দেখতে পছন্দ করি।

Chunk 2 : এছাড়াও আইফেল টাওয়ারের আশেপাশে কিছু ভাল রেস্তোরাঁ এবং হোটেলের সাথে কিছু পর্যটন স্থানের পরামর্শ দিন যা আমি ঘুরে দেখতে পারি?

Figure 7: Examples of complete cross-domain, multi-intent user utterance. The user utterance is chunked into respective segments as per the proposed methodology

Table 1: Top 5 Similarity Scores for k values per chunk

k per Chunk	Chunk 1	Chunk 2	Similarity
k_1	8	8	76.03
k_2	8	16	76.03
k_3	8	12	70.48
k_4	8	32	70.48
k_5	12	8	62.93

User Utterance	k=8	k=12	k=16	k=32
আপনি কি আমাকে প্যারিস যাওয়ার জন্য একটি ফ্লাইটের পরামর্শ দিতে পারেন যেহেতু আমি খুব ঘন ঘন ভ্রমণকারী এবং বিশ্বের বিভিন্ন স্থান ঘুরে দেখতে পছন্দ করি। এছাড়াও আইফেল টাওয়ারের আশেপাশে কিছু ভাল রেস্তোরাঁ এবং হোটেলের সাথে কিছু পর্যটন স্থানের পরামর্শ দিন যা আমি ঘুরে দেখতে পারি?	{'Intent': ['flight_booking', 'travel_recommendation'], 'Domain': ['travel', 'recommendation']}	{'Intent': ['flight_booking', 'travel_recommendation'], 'Domain': ['travel', 'recommendation']}	{'intent': ['recommendation_locations', 'recommendation_movies'], 'domain': ['recommendation']}	{'intent': ['transport_query', 'recommendation_locations'], 'domain': ['transport', 'recommendation']}
	{'intent': ['restaurant_recommendation', 'hotel_recommendation', 'tourism_recommendation'], 'domain': ['recommendation', 'restaurant', 'hotel', 'tourism']}	{'intent': ['recommendation_locations'], 'domain': ['recommendation']}	{'intent': ['recommendation_locations'], 'domain': ['recommendation']}	{'intent': ['recommendation_locations'], 'domain': ['recommendation']}

Figure 8: Results for different k values where k = 8, 12, 16, 32, for different chunks using RAG and COT from GPT-3.5 Turbo

User Utterance	k=8	k=12	k=16	k=32
আপনি কি আমাকে প্যারিস যাওয়ার জন্য একটি ফ্লাইটের পরামর্শ দিতে পারেন যেহেতু আমি খুব ঘন ঘন ভ্রমণকারী এবং বিশ্বের বিভিন্ন স্থান ঘুরে দেখতে পছন্দ করি। এছাড়াও আইফেল টাওয়ারের আশেপাশে কিছু ভাল রেস্তোরাঁ এবং হোটেলের সাথে কিছু পর্যটন স্থানের পরামর্শ দিন যা আমি ঘুরে দেখতে পারি?	{'intent': ['recommendation_locations', 'recommendation_movies', 'transport_query'], 'domain': ['recommendation', 'transport']}	{'intent': ['recommendation_locations'], 'domain': ['recommendation']}	{'intent': ['flight_booking', 'hotel_booking', 'restaurant_recommendation'], 'domain': ['travel', 'hospitality', 'food']}	{'intent': ['flight_booking', 'hotel_booking', 'restaurant_recommendation'], 'domain': ['travel', 'hospitality', 'food']}

Figure 9: Results for different k values where k = 8, 12, 16, 32, for a complete sentence using RAG and COT from GPT-3.5 Turbo

6 Conclusion

This study presents the concept of cross-domain multi-intent identification for low-resource languages by utilizing a pivot high-resource language. We have conducted our experiments on Indian regional languages. We utilized segment-based retrieval augmented generation by employing a vector database in conjunction with chain of thought prompting, resulting in enhanced accuracy and efficiency. The embeddings for languages with abundant resources were put in the vector database to facilitate efficient retrieval and indexing. We evaluated the necessity of splitting a complex sentence that spans many domains into segments and performing RAG (Retrieval-Augmented Generation) with various k values. Here, k refers to a few-shot examples in a high-resource language which are semantically correlated, and these examples are used to build contexts for prompts used in Chain of Thought Prompting. Future research could primarily concentrate on extending this study to various NLP tasks and multiple low-resource endangered unknown languages. Additionally, it should prioritize the exploration of various ways to identify the k most ideal values for few-shot prompting in low-resource languages.

References

Bo An. 2023. Prompt-based for low-resource tibetan text classification. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 22(8).

Sourojit Ghosh and Aylin Caliskan. 2023. Chatgpt perpetuates gender bias in machine translation and ignores non-gendered pronouns: Findings across bengali and five other low-resource languages. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, page 901–912, New York, NY, USA. Association for Computing Machinery.

Nuno M. Guerreiro, Duarte Alves, Jonas Waldendorf, Barry Haddow, Alexandra Birch, Pierre Colombo, and André F. T. Martins. 2023. Hallucinations in large multilingual translation models.

Ayiguli Halike, Aishan Wumaier, and Tuergen Yibulayin. 2023. Zero-shot relation triple extraction with prompts for low-resource languages. *Applied Sciences*, 13(7).

Haoyang Huang, Tianyi Tang, Dongdong Zhang, Wayne Xin Zhao, Ting Song, Yan Xia, and Furu Wei. 2023. Not all languages are created equal in llms: Improving multilingual capability by cross-lingual-thought prompting.

Feihu Jin, Jinliang Lu, Jiajun Zhang, and Chengqing Zong. 2022. Instance-aware prompt learning for language understanding and generation.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. 55(9).

Yao Lu, Max Bartolo, Alastair Moore, Sebastian Riedel, and Pontus Stenetorp. 2022. Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity.

Ercong Nie, Sheng Liang, Helmut Schmid, and Hinrich Schütze. 2023. Cross-lingual retrieval augmented prompt for low-resource languages. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8320–8340, Toronto, Canada. Association for Computational Linguistics.

Soham Parikh, Quaizar Vohra, Prashil Tumbade, and Mitul Tiwari. 2023. Exploring zero and few-shot techniques for intent classification.

Nikolaos Stylianou, Despoina Chatzakou, Theodora Tsikrika, Stefanos Vrochidis, and Ioannis Kompatsiaris. 2023. Domain-aligned data augmentation for low-resource and imbalanced text classification. In *Advances in Information Retrieval*, pages 172–187, Cham. Springer Nature Switzerland.

Yufei Wang, Can Xu, Qingfeng Sun, Huang Hu, Chongyang Tao, Xiubo Geng, and Daxin Jiang. 2022. PromDA: Prompt-based data augmentation for low-resource NLU tasks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4242–4255, Dublin, Ireland. Association for Computational Linguistics.