Cross-Lingual Intent Classification using BERT: A Multilingual Approach

Abstract:

With conversational AI taking center stage in the world, systems such as chatbots, conversation commerce, virtual assistants and customer support are required to function across multiple languages. This makes multilingual intent classification a critical task. This project explores the use of transformer model Bidirectional Encoder Representations from Transformers (BERT) for intent classification across various languages. The project aim is to classify user intents using pre-trained multilingual BERT (mBERT) in a multiple language dataset which includes English, Spanish, French, and Hindi.

The main objective is to evaluate the performance of mBERT in recognizing and classifying user intents in a multi-lingual context for which challenges such as language diversity, date sparsity, and domain-specific vocabulary need to be addressed. The aim is to build a robust model that can improve accessibility and accuracy of AI systems for global users by fine-tuning BERT on a multilingual dataset. We will evaluate the model's performance using various metrics such as accuracy, precision, recall, and F1-score, across different languages.

This project will contribute to the development of more inclusive and adaptable conversational AI models which can handle diverse language inputs thereby paving the way for seamless cross cultural interactions in AI applications.