Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)
Doctoral Consortium Track

Exploring Multilingual Intent Dynamics and Applications

Ankan Mullick
Computer Science and Engineering Department, IIT Kharagpur, India
ankanm@kgpian.iitkgp.ac.in

Abstract

Multilingual Intent Detection and explore its different characteristics are major field of study for last few years. But, detection of intention dynamics from text or voice, especially in the Indian multilingual contexts, is a challenging task as there are many languages spoken together in one context and multilingual syntaxes appear interleaved in a single conversation. So, my first research question is on intent detection and then I work on the application in Indian Multilingual Healthcare scenario.

Speech dialogue systems are designed by a predefined set of intents to perform user specified tasks. Newer intentions may surface over time that call for retraining. However, the newer intents may not be explicitly announced and need to be inferred dynamically. Hence, here are two crucial jobs: (a) recognizing newly emergent intents; and (b) annotating the data of the new intents in order to effectively retrain the underlying classifier. The tasks become specially challenging when a large number of new intents emerge simultaneously and there is a limited budget of manual annotation. We develop MNID (Multiple Novel Intent Detection), a cluster based framework that can identify multiple novel intents while optimized human annotation cost. Empirical findings on numerous benchmark datasets (of varying sizes) show that MNID surpasses the baseline approaches in terms of accuracy and F1-score by wisely allocating the budget for annotation. We apply intent detection approach on different domains in Indian multilingual scenarios - healthcare, finance etc. The creation of advanced NLU healthcare systems is threatened by the lack of data and technology constraints for resource-poor languages in developing nations like India. We evaluate the current state of several cutting-edge language models used in the healthcare with the goal of detecting query intents and corresponding entities. We conduct comprehensive trials on a number of models different realistic contexts, and we investigate the practical relevance depending on budget and the availability of data on English.

1 Research Questions and Contributions

In my PhD, in the last three years, I have addressed the following Research Questions and exploring further -

RQ1: Exploring Multiple Novel Intent Detection: To formally describe the problem setting, let there be a dataset W containing overall N classes. However, the value of N is not known apriori. Let $T \in W$ be the test set and $W - T = D$ be the rest of the dataset, out of which $|D_{init}|$ ($<< |D|$) labelled data of N_{init} ($< N$) classes is initially provided, while the rest of the data is unlabelled. The task is to design an algorithm to (a). detect all the remaining $N - N_{init}$ classes and (b). spent a limited budget (B - $|D_{init}|$) to annotate high fidelitous new datapoints, so that the classifier can achieve high accuracy when retraining.

Contribution: RQ1 - Solution Overview [Mullick *et al.*, 2022]: The solution steps are as follows: (a) Identify the OOD (out of distribution) datapoints using Few-shot OOD (FS-OOD) which do not belong to the initial classes. This can be considered as a preprocessing step. (b) Use a part of the allotted budget to annotate a portion of these OOD datapoints. These points (for annotation) are selected by repeatedly running K-Means clustering algorithm with increasing number of clusters as input, and choosing cluster centre points to identify the unknown classes. **Rationale:** The intuition is that each cluster hosts a separate intent, hence annotating the cluster centres would lead to discovery of maximum number of novel intents. (c) Further identify the classes which are well clustered (**Rationale:** good cluster with a single intent) in feature space and which are not (bad clusters with multiple intents). Use another portion of the budget to increase the annotations of not-so well formed clusters to build up a classifier. (d) Use the classifier to classify points from the clusters. Identify low-confidence points (most uncertain) from the bad clusters and annotate them (gold annotation). High-confidence points from good clusters are silver annotated. (e). Retrain the classifier.

RQ2: Application in Multilingual Healthcare Scenario:

Every nation places a high priority on healthcare. Millions of health-related questions are asked by people worldwide in an effort to acquire an answer from a subject matter expert. Most of these inquiries are about the patients' medical histories, potential drug interactions, worries about diseases, treatment regimens, and similar topics. By facilitating the spread of important information, diagolue systems for healthcare play a crucial role. Yet, a lack of useful data is the main barrier to the development of such solutions for low-resource languages. We explore in Indian contexts. India is a country with a diverse language speaking population suffering from abject poverty and low-economic status. Developing automatic healthcare systems in India is undoubtedly difficult due to the country's linguistic diversity and complex socioeconomic environment, and there aren't many linguistic healthcare resources. Massively Multilingual Transformer based Language Models (MMLM) have made remarkable progress on a variety of downstream uses in order to overcome this language barrier. However, the practical effects of these developments in the Indian healthcare system are still unexplored.

Contribution RQ2 - Solution Overview [Mullick *et al.*, 2023]: The contributions are five folds: (1) Develop Indian healthcare datasets with intent and entity labelled. We propose IHQID-WebMD and IHQID-1mg (annotated by domain-experts) comprising of frequently asked questions from users. (2) Healthcare query intent (4 types - disease, drug, treatment and other) detection and corresponding entity extraction (3) Although the large language models have demonstrated their efficacy in nearly all NLU operations, we want to assess how well they perform in identifying the intent and corresponding entities for realistic domain-specific healthcare scenarios in the Indian context. We analyze how should we prioritize the research and resource building investments for the economically backward countries with a high percentage of multilingual population - availability of English training data (cost effective) vs multilingual training data (costly). Keeping this in mind, all our experiments have been carried out using both monolingual and multilingual setups of these models to point out the best possible language models and techniques. (4) Through extensive experiments, we recommend to use back-translation of test queries to English in real-life scenarios as a reasonable choice when we have access to English training data. However, with sufficient budget, the same strategy can be applied to both train and test queries in target languages. (5) Our findings imply that the back-translation of queries using an intermediate bridge language (Like - Hindi) proves to be a useful strategy in the intent recognition experiments for low resource languages (like - Bihari) close to the bridge language.

The evaluations are based on two possible real-life scenarios: 1) **Scenario A:** access to only English training data (less costly) and in 2) **Scenario B:** access to manually written training queries in all the target languages (very expensive). During inference/testing, we expect all the queries are in the corresponding target languages. There are three different setups for Scenrio A. *Setup 1) Backtranslated Test (S1): [Translate-Test]* Training models on the English queries, and evaluate the intent detection and entity extraction tasks in different languages by automatically backtranslating the test queries into English. *Setup 2) Zero-Shot Cross-Lingual Test (S2):* Training on the English data (scraped from WebMD and 1mg) and use it for inference on test queries in Indic languages. *Setup 3) Bridge Language Backtranslation (S3):* Here a relatively low-resource language is first translated to an intermediate bridge language ('Hindi') and then finally to English. There are two setups in Scenario B - *Setup 4) Train-Test on Indic Data (S4):* Training dataset in indic languages to train NLU models in different target languages. *Setup 5) Full Backtranslation (S5):* In this setup, both train and test data are backtranslated to English. This is useful for the countries with poor technical setups for low-resource languages. In all back translation experiments, we use Bing Api [1].

2 Conclusion and Future Direction

A) [RQ1] In the near future I would like to work on several limitations of MNID approach - 1) Our system is unable to detect intents where classes are closely similar to each other. For example, "I topped up but the app declined it" query in Banking dataset is from "top up reverted" category but our system detects as "top up failed" intent as both intents are similar. 2) Our system fails to identify intent periodicity (Appear/disappear) in continuous streams.

B) [RQ2] I am currently exploring following limitations - issues in mis-classification due to model prediction error. For an example, 'How common is syphilis' is of 'disease' intent category but model wrongly predicts it as 'other' category. Another reason is the mis-classification due to incorrect translation of the medical entities such as the disease *'uticartia'* has been transformed into *'ambat'* during backtranslation. I am also working on following directions - a) Heath Query Analysis in Indian Multilingual Speech Contexts with different dialects. b) Generate appropriate responses for English and Indic languages for different healthcare queries.

C) [RQ3] **Dialogue System for HealthCare**: To build such an efficient system, I am currently working on developing - sizable amount of interlinked multilingual question and answer pairs (QA), and a knowledge graph that supports multilingual entity and relationships. For the multilingual QA training data, we plan to generate a handful of such real multilingual QAs from real monolingual resource heavy (let's say English) labelled QA dataset by using Google or Bing translation APIs. Then, the framework is to mask some of the tokens and asking to predict those masked tokens in the desired language. The presence of a large amount of training data will help to build the end-to-end framework given the natural language question (NLQ) in multilingual dialect.

Acknowledgments

I am Thankful to Prime Minister Research Fellow (PMRF) Grant for supporting my PhD. I am thankful to PMRF, Micrososft, Google, ACM, AAAI for providing different grants.

References

[Mullick *et al.*, 2022] Ankan Mullick, S Purkayastha, Pawan Goyal, and Niloy Ganguly. A framework to generate high-quality datapoints for multiple novel intent detection. In *Findings of the ACL: NAACL*, 2022.

[Mullick *et al.*, 2023] Ankan Mullick, I Mondal, S Ray, R Raghav, G Chaitanya, and Pawan Goyal. Intent identification and entity extraction for healthcare queries in indic languages. In *Findings of EACL*, pages 1825–1836, 2023.

[1] https://www.microsoft.com/en-us/translator/business/translator-api/