

Bank of Baroda Hackathon - 2022

Team : Alin Wehsac

Team Bio:

Date: 15/09/2022

TIMES INTERNET

//01



TECHGIG

Problem Statement

Contact Center is a very important touch point for customer service. Customers interact with the contact center for a variety of services fulfillment, information about products and services, complaints, feedback, and interactions. For enhancing the contact center service levels, the customer interactions need to be analyzed and need to be mapped with customer satisfaction. Banks want to automate the process with use of AI based techniques to measure and monitor various KPIs such as call data quality, customer sentiment, Call Hygiene, customer satisfaction etc.

→ **Why did you decide to solve this Problem statement?**

We were intrigued by the problem presented, we always had a keen interest in audio signal processing and this was a good opportunity not only to explore new domains but implement a working prototype. We formulated our solution so as to make use of SOTA Deep Learning practices and techniques to achieve highly accurate results.

User Segment & Pain Points

→ Which user / advertiser segment would be early adopter of your product & why?

The major part of the **user segment** would be all banks in general, and any other businesses that possess a functional call centre and require call centre analytics. The top industry segments that would utilize such technology would be inclusive of

- Healthcare
- Automotive
- Telecommunication
- Banking & Financial Services
- Information Technology Services
- Travel and Transportation

In addition to this, as early adopters are usually less concerned with price and risk, our user segment would even comprise of startups looking for a easy to use call analytics platform with a KPI dashboard displaying all the relevant information they would need.

Pain points

- A customer calling with a bad quality of audio, as the **representative** wouldn't be able to understand what the customer is saying, in addition, such audio samples would contain high amount of noise which would make the preprocessing / noise removal a behemoth task.
- A customer using **NSFW** language (**call hygiene**), performing a profanity check would automatically get rid of such audio recordings.
- Lack of appropriate skill, required to operate the call center analytics tool. Although the solution proposes the creation of an “**easy to use**” tool, it would still be necessary to educate the employee on its proper usage technique.
- It would be hard to provide a accurate sentiment analysis on audio samples consisting of fluctuating customer emotions.

Pre-Requisites

- ➔ **What are the alternatives / competitive products for the problem you are solving?**
- There would be several alternatives to choose from, as there are many services providing a real time KPI monitoring dashboard for all kinds of businesses.
- Competitive products could comprise of solutions that use an API for conversion of call audio Speech to Text (STT), or some other SOTA model that gives a high accuracy on STT, which is also the evaluation metric for the problem presented. In combination with this STT module, a highly accurate sentiment analysis model can be used which finally gives an accurate sentiment analysis. Such an approach can't be integrated into our solution mainly due to lack of hardware required for computation.

Azure tools/resources

- ➔ **Azure tools or resources which are likely to be used by you for the prototype, if your idea gets selected**
- Linux Virtual Machines with Hardware Acceleration
(for training the model and hosting the dashboard)
- Azure Database for MySQL & Table Storage for storing various data

Supporting Functional Documents

- The approach chosen is that of Deep Learning. DL methods are used for both stt (Speech to Text) and for performing sentiment analysis on the text extracted from the audio samples.

The model used for Multilingual stt is XLSR which stands for Cross Lingual Speech Representation, refers to the model's ability to learn speech representations that are useful across multiple languages, derived from the wave2vec2 model developed by facebook AI. XLS-R learns contextualized speech representations by randomly masking feature vectors before passing them to a transformer network during self-supervised pre-training.

- The text data extracted from the call audio samples is used for performing sentiment analysis. We use a SOTA technique, BERT for achieving this. The model can be fine tuned for multiple languages (Multilingual).

- BERT is a transformers model pre trained on a large corpus of multilingual data in a self-supervised fashion.

More precisely, it was pretrained with two objectives:

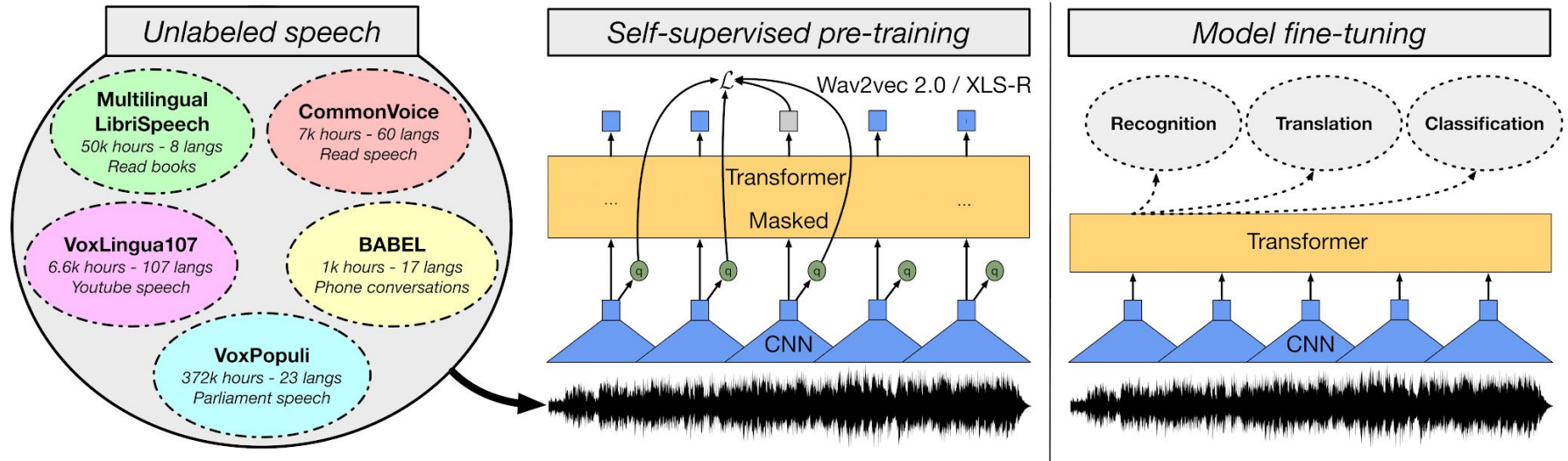
- Masked language modeling (MLM): taking a sentence, the model randomly masks 15% of the words in the input then run the entire masked sentence through the model and has to predict the masked words. This is different from traditional recurrent neural networks (RNNs) that usually see the words one after the other, or from autoregressive models like GPT which internally mask the future tokens. It allows the model to learn a bidirectional representation of the sentence.
- Next sentence prediction (NSP): the models concatenates two masked sentences as inputs during pretraining. Sometimes they correspond to sentences that were next to each other in the original text, sometimes not. The model then has to predict if the two sentences were following each other or not.

Key Differentiators

- ➔ **How is your solution better than alternatives and how do you plan to build adoption?**
 - We are using XLSR for stt, which is pretrained with the WER (Word Error Rate) metric, so the model naturally performs well in accordance to our current problem statement.
 - And for sentiment analysis, we are using transformers, **BERT**, which uses multilingual transformers and is highly accurate.
 - This gives us two modules which will be integrated together to create a working prototype.

Supporting Diagrams

Diagrammatic representation of the XLSR:



TECHGIG

Thank You

Vishal Lokare

Ashutosh Rai