CLAP - Speech to Text

R&D Project Report
submitted in partial fulfillment of the
requirements for the degree of
Master of Technology
by

Vamshi (183050026), Sahil (183059002), Smartika (17305T001)

under the guidance of

Prof. Kameswari Chebrolu

Department of Computer Science and Technology
Indian Institute of Technology, Bombay
Mumbai 400 076

Contents

List of Figures

Chapter 1

Introduction

1.1 Problem Statement

1.2 Outline of Report

Chapter 1 of the report gives a brief introduction and motivation behind creating crowdsourced system to create rich data-set of Indian languages. Chapter 2 focuses on the background work which involves details of the existing system, its architecture and loopholes. Chapter 3 defines the features added to the already existing system and their implementation. Chapter 4 emphasizes on the future work which can be taken up.

Chapter 2

Background Work

The solution proposed for the problem statement mentioned in previous chapter is a crowdsourced platform for collecting rich audio data for Indian languages. The architecture of the system is such that users needs to register using the Android app and select languages they are comfortable speaking. The speak as well as verify tasks are then allocated to the user. In order to retain users, incentives in the form of certificates and t-shirts are provided. Below section discusses about the technical details of the solution.

2.1 Proposed System

The system proposed comprises of a web server, an Android application and a dashboard. The web server is build in Dja go, a web development framework that uses Model View Controller(MVC) architecture. The workflow involves the following steps:

Tasks creation: The task creation for different languages like Hindi, Marathi, Tamil etc. has the same logic. The task creation logic opens a text file corresponding to the particular language. It then reads the text file line by line, each line becoming a separate task in the database. There is no limit on the maximum number of words in that a task can include. The minimum number of words that a task can include is 5. The information stored for each task consists of a unique identifier(by concatenating the primary key with 'File'), the text of the task and the language the task belongs to.

Tasks allocation: There exists two types of tasks, namely speak tasks and verify tasks. Initially, speak tasks are allocated to registered users based on the language selected by them. After the speak tasks are completed, verify tasks get

created which are also allocated similarly. The speak tasks involves reading a sentence and recording it correctly. The verify tasks involves listening to a speech recording and matching it with the text. All speak and verify tasks were allocated to atleast 3 users(configurable).

User performance: The ranking of user is evaluated based on the total tasks(speak and verify) completed. Based on the ranking, appropriate levels are allocated to the user. The top users on leader board are given t-shirts and other prizes.

Dashboard: The dashboard lets the admin of the system monitor various statistics like language-wise tasks, daily/monthly/yearly number of tasks, pending tasks, user profile.

2.2 System Experiment Feedback

Internal trials were conducted for the system. During those trials, various shortcomings came up which were expected to be supported by the system and are listed below:

1. The task creation logic just read the source text file line by line making each line into a task.

2. The task creation logic didn't take into account the different delimiters in different languages. For Ex: In Hindi, 'matra' was not taken into consideration while creating the tasks.

3. The system did not put a limit on the maximum number of words that can be in a particular task. This resulted in tasks having very large number of words. There was only a limit on the minimum number of words that could be in one task which was kept same for all the tasks.

4. All the tasks were considered the same. There was no logic to distinguish between "conversational" , "dialogue" and "regular" tasks.

5. The system didn't record which tasks came from which book.

6. There was no proper organization of books on the server. All the books of the same language were kept in the same folder irrespective of their source.

7. The tasks allocation was random so it was difficult to track the number of tasks for which all audios were collected or verified.

8. The audios were stored in a common folder which made it difficult to differentiate between multiple language audios.
9. The tasks allocation logic allowed diversity in language only for first time allocation for the user. For all other times, the allocation was sequentially.
10. The text to be recorded in speak task was sometimes incorrect. Even then the task was allocated as speak as well as verify tasks.
11. The app allowed submitting the verify tasks without giving feedback on audio.
12. The user levels involved completion of very few tasks.
13. In speak task, provision for reporting additional comments like too long audio, grammatical mistakes, etc was not there.

Chapter 3

Improvements in the system

As mentioned in Chapter 2, experiments were conducted and various shortcomings were found in the existing system. Below section discusses about features to be added to overcome the shortcomings.

3.1 Features Added

From the feedback obtained, the features to be added were analyzed from the perspective of users. The following features were identified to be added:

1. Task creation mechanism which takes into consideration different delimiters and avoid parsing incorrect sentences as task.
2. Store additional details about the text: source, task type, etc.
3. Proper folder structure to store the books.
4. Better task allocation mechanism that keeps track of the number of audios verified and better storage of the audio files.
5. To facilitate data-set creation, csv should be maintained to store the appropriate details.
6. Structure user performance and proper levels assignment based on the number of tasks completed.

3.2 Implementation

This section discusses about the implementation of the above mentioned features.

3.2.1 PDF Conversion

As there were no Unicode text files available, the conversion of available PDF to text files has been done. This has been done with the use of an OCR engine(Tesseract in this case). This process has been done for Marathi, Malayalam, Telugu and Tamil languages. Tesseract needs "traineddata" files for each language for conversion. There are two kinds of traineddata files: fast and best. As "best" traineddata files are slow but have higher accuracy, they were used for conversion.

3.2.2 Task Creation

The current logic of task creation was modified to take into account various delimiters in different languages. Also three types of tasks namely 'conversational', 'dialogue' and 'regular' were generated instead of clubbing them all into one category. The conversational tasks includes those within double or either single quotes. The dialogue category includes tasks which one would normally see in a screenplay script. The regular tasks consists of regular sentences delimited by either a full stop(.), question mark(?) or an exclamation mark(!). To implement this functionality, regex in python is used. Three different kinds of regex(one for each category) is used to extract out the tasks of relevant category. Also the logic has been modified to put limit on the maximum number of words that a task can contain. The number of minimum and maximum words that a task can contain is different for different languages. The schema of the database is also modified to include the book name and the type of task. As the corpus for the languages was not available, an OCR engine was used to convert PDFs into their corresponding text files. As OCR engine can introduce various errors(weird characters or extra spaces,tabs etc.), the task creation logic has been modified to remove those errors(In some case remove the sentences containing those errors).
Some of the tasks created for various languages are shown below:

LANGUAGE	TEXT
Marathi	थोडावेळ इकडचे तिकडचे बोलणे झाल्यावर सरपंचांनी मुख्य विषयाला हात घातला.
Marathi	या शिवाय आणखी काय काय खर्च लागेल.
Marathi	परवाच जवळपास पस्तीस फुटावर पाण्याचे सात मोठ्ट जोरदार झरे तिथे लागले होते.
Marathi	हीच आपली त्रिकाल संध्या आहे.
Marathi	हे माझे आपणा सर्वांच्या आणि सोमजाईच्या साक्षीने दिलेले वचन आहे.
Marathi	या गावात आल्यावर माझी झोळी कम् शबनम मी मारुतीच्या देवळात लावली आहे.
Marathi	एकीकडे डोक्यात विचार चालू असताना ते फुलझाडांना पाणी घालण्याचे कामही करीत होते.
Marathi	चारही बाजुन ओपन असल्यामुळे छान हवा आत येत होती.

Figure 3.1: Marathi

Tamil	மீண்டும் மீண்டும் முகிலன் இந்தக் கேள்வியைக் கேட்டுக் கொண்டிருந்ததால், முகிலனின் அப்பா இளவெழில் வாயைத் திறந்தார்.
Tamil	கண் விழித்துப் பார்கையில் சுருள் சுருளாக ஏதோ தெரிந்தது.
Tamil	தப்பு செய்யுறவன பத்தி ஒரு அடையாளமும் தெரியாது.
Tamil	இத்துனூண்டு பூச்சிய தூக்கிக் கொண்டு போய் புத்துக்குள்ள வைக்கிறதுக்கு கூட நம்ம உடம்புல தெம்பில்ல.
Tamil	அப்ப நீ நினச்ச இன்ச்டிட்யுசன்ா ப ாா கிடைக்கும்மா பா?
Tamil	உனக்கு அப்புடி எதுவும் ஆகக்கூடாதுன்னு தான் சொல்றேன்.
Tamil	அந்தப் பெண்ணையும் காணவில்லை!
Tamil	கவிதைகளிலும் புதிய முயற்சிகள் கையாளப்பட்டிருக்கிறது.
Tamil	நம்ம வீட்டுக்கு போய் பேசலாங்க.

Figure 3.2: Tamil

LANGUAGE	TEXT
Hindi	मधुमक्खियाँ उनसे रस जमा करती हैं
Hindi	ज़मीन नरम और नम हो गई थी और उसमें से सुगन्ध आ रही थी
Hindi	कच्चे केले की सब्ज़ी बनाकर खायी जाती है
Hindi	समुद्र के पास के इलाकों में नारियल के पेड़ बहुतायत से दिखते हैं
Hindi	लेकिन बागों में लगे आम के पेड़ तो ३मीटर तक के ही होते हैं
Hindi	रेशम के कीड़े भी बेर के पत्ते खाते हैं
Hindi	अमरूद की छाल से रेशम को रंगने के लिए रंग बनाया जाता है
Hindi	संतरे की पत्तियाँ गहरी हरी और चमकीली होती हैं

Figure 3.3: Hindi

Telugu	అంటూ తలువులు మూసి గడియ వేసుకోవటం వినిపించేసరికి కళ్ళనీళ్ళు తిరిగిపోయాయి.
Telugu	ఇది బంగారపు గొలుసు కాదమ్మా!
Telugu	తిరిగి కాసేపు హీరో హీరోయిన్ ల నిలువెత్తు కటౌట్ లపైపు చూసింది.
Telugu	ముగ్గురూ ఒకరి ప్రక్క మరొకరుగా నడుస్తున్నారు.
Telugu	స్వాధీనంలోకి రాణి కాళ్ళను ఈడ్చుకుంటూ లేని ఓపికను తెచ్చుకుని నడవసాగింది.
Telugu	మీరు కిరాయికి బండితోలుకుని బ్రతికేవారన్నది.
Telugu	కాసేపే జాకెట్టులోనుంచి డబ్బుకట్టతీసి అందులో నుంచి నాలుగు యాభై కాగితాలు తీసి ఇచ్చింది.
Telugu	నడుస్తుంటే పిరుదులుని తాకే పి వాలుజడ!

Figure 3.4: Telugu

The sentences generated by regex are passed by two filters: One which checks if they fulfill the number of words and second condition is to ensure that they do not contain any weird characters. Below is a table showing the wastage of tasks due to the two filters.(All values are in percentage)

Marathi

Task type	Length Limit	Removing weird sentences	Both	Total
Regular	39.04	7.5	6.1	52.64
Conversational	36.1	4.4	40.3	80.8
Dialogue	35.1	11.61	15.03	61.74

Tamil

Task type	Length Limit	Removing weird sentences	Both	Total
Regular	35.93	7.7	5.04	48.67
Conversational	38.5	3.1	33.1	74.7
Dialogue	24.93	15.5	22.07	62.5

Hindi

Task type	Length Limit	Removing weird sentences	Both	Total
Regular	24.7	14.56	14.14	53.4
Conversational	20.1	14.5	32	66.6
Dialogue	29.4	13.1	26.5	69

Telugu

Task type	Length Limit	Removing weird sentences	Both	Total
Regular	53.03	2.1	0.7	55.83
Conversational	44.39	1.7	19.6	65.69
Dialogue	43.8	13.04	14.98	71.82

As one can see the wastage is mostly due to the length limits. If one increases the maximum number of words that can be in a task the wastage drops considerably.

Not all books contribute equally to the task generation process. Some books are huge and generate more tasks while others generate similar number of tasks. Here is the graph of each language showing the number of tasks generated by each book.

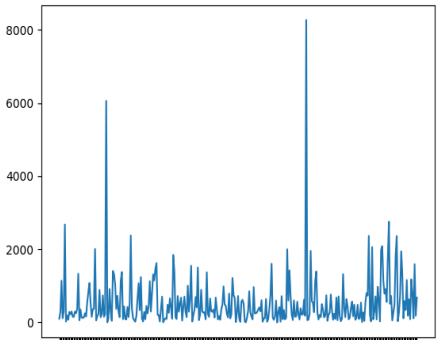

Figure 3.5: Marathi Tasks Distribution

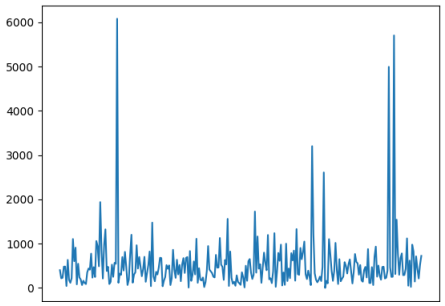

Figure 3.6: Tamil Tasks Distribution

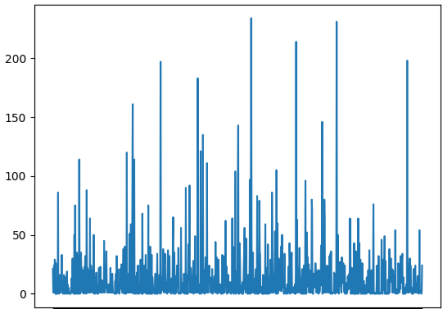

Figure 3.7: Hindi Tasks Distribution

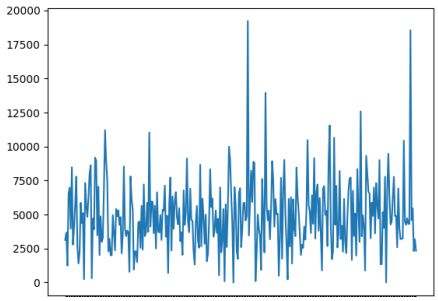

Figure 3.8: Telugu Tasks Distribution

3.2.3 Tasks Allocation

Speak Tasks

In order to allow tasks to be completed end to end, the current logic of task allocation was modified. The new approach takes a batch of tasks of configurable size and allocates them to the users. The maximum number of users to which a task will be allocated is also kept configurable. The first set of allocation is of tasks which have no assigned users. The second set is of tasks which have one assigned user and so on. Once the batch is exhausted, the next batch is picked. Figure 3.9 illustrates one such batch allocation.

Figure 3.9: Batchwise Allocation

Once the audio has been recorded, the user is prompted to give feedback on the text. Figure 3.10 shows the app screenshot for the same. The text files that receive negative feedback are not further allocated as speak tasks.

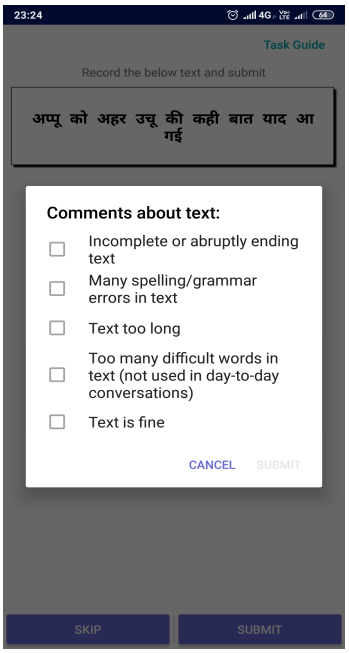

Figure 3.10: Text Feedback

A csv file is maintained for all the audio files submitted. The csv maps the audio names with task ids and stores additional information. Figure 3.11 shows the sample updated csv file.

Audio File Name	Text	Task_Id	Language	Source	Sentence Type
66034_1	सबसे ज्यादा महत्वपूर्ण बात यह है कि काली मिर्च सबसे ज्यादा कीमती पदार्थ था	22049	Hindi	Pratham	Regular
66207_1	बरोबर दसऱ्यालाही बुढ़वितो आहे	22051	Marathi	esahity	Dialouge
66206_1	सन्तु सर्वे सन्तु निरामय	22050	Marathi	esahity	Dialouge
66035_1	' यह पौधा दुनिया के अन्य बहुत से भागों में फैल गया	22052	Hindi	Pratham	Regular
66037_1	दक्षिणी अमेरिका ग्लोब के दूसरी तरफ़ मिला	22054	Hindi	Pratham	Regular

Figure 3.11: CSV with audio and task details

Verify Tasks

Once the speak tasks have been completed, verify task is created. Since the tasks allocated to users is a combination of speak and verify tasks, it is ensured that

there is diversity in tasks assigned. Similar to the speak tasks, verification logic has been changed. The verification tasks are now assigned in batches. For each verification task there is feedback given by the user, about the audio quality of the task. Each verification task is assigned to 2 users and their feedback of audio quality is checked. If both users give different feedback, then this verification task will be assigned to a 3rd user otherwise it is not allocated further. Hence, any verification task is assigned to atmost 3(configurable) users. Figure 3.12 shows the app screenshot where user is asked to give feedback about the audio quality.

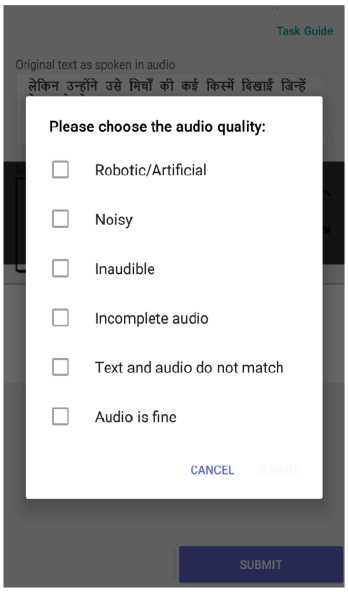

Figure 3.12: Verification Feedback

3.2.4 Android App related changes

Speak Task Feedback

In the earlier version of the android app, there was no provision in speak task, for reporting feedback about the task text. Now, we have added this into the app, during the speak task submission, user can select comments(feedback) for the task. These comments are very useful to categorize the tasks. For an example, if a user gives negative feedback, these tasks can be removed or should not be assigned to new users. Figure 3.10 shows the app screenshot for the speak task feedback as implemented in the app.

Verification task Feedback

The older version of app allowed submitting the verification tasks, without giving feedback on audio. We have changed this and now we mandate user to give feedback on the audio quality. The user cannot submit the task without giving his feedback.

We have also moved the verification feedback option from verification task screen to alert dialogue. In this way, it is easy for user to always give feedback without forgetting. Figure 3.12 shows the app screenshot for the verification feedback.

Removed filter for tasks

The older version of app had a filter that could have been used, to filter the tasks based on type, language, etc. Since only 5 tasks are there with a user at any time, we didn't the filter to be much useful. So we have removed that filter.

Tutorial updated

We have updated the tutorial to support with the latest app features. The tutorial has been simplified by keeping minimal text. Since, our app now only supports speak and verification tasks, tutorial pages corresponding to label task were removed.

Profile section

The profile section has been updated, rating has been replaced by tasks done. We also removed statistics related to label tasks.

There were few other minor changes that are made in the android app. The offers section has been updated, performance summary and Leader board sections are also updated.

Chapter 4

Future Work