Conversational Question Answering System: (Add value to the system by introducing interaction summarizer and sentiment analyzer)

[By Bhavesh Kotwani (2018B1A70614P)]

1. Introduction

1.1 Background

With the ever growing market needs, services like customer care, HR processes, Healthcare and features like autocomplete in computer software are gaining importance, and these industries cannot depend on human labor alone, and therefore Conversation AI comes into picture. Companies employ chatbots and voice assistants to provide the needed services to their customers, and these services are performed omnichannel by major companies, and all this has been achieved with the application of Conversational AI.

How does Conversational AI work?

Conversational AI, as the name suggests uses Machine Learning for understanding human conversations. The basic infrastructure of any chatbot or voice assistant includes an NLU Framework (Natural Language UnderstandingFramework), which is like the library of the virtual assistant. The NLU framework contains all the examples of the user responses categorized as intents, along with specific keywords (entities) which are stored and processed by the assistant. These examples of human conversations are used to train the chatbot for any possible conversation a customer might come up with.

For practical purposes, NLU files can be really bulky since with time, the examples might increase to improve the accuracy of the virtual assistant. For this purpose, companies store these files in specific servers(or cloud), and provide an endpoint to the chatbot for the purposes of referencing so that the efficiency is not dropped.

The Dialog Framework within the virtual assistant is responsible for holding conversations with the customer/user. The bot first understands the intent of the user using the NLU Framework and the trained Machine Learning Model and then performs a subsequent action based on the predefined stories or rules of the conversation. Developers can define a set of rules which would make a flow of conversation between the bot and the user, and help achieve the desired tasks. One can define any set of tasks like payment, billing, normal conversations, reminders, etc. in various domains of businesses increasing the cost efficiency, managing sales and customer engagement and providing the companies with the much wanted scalability.

1.2 Motivation

We are working towards achieving a multilingual conversational bot functionality, which can understand the human sentiment and summarize each conversation from the user, while also performing context switching with great efficiency. While there has been considerable progress in terms of the tasks which could be performed using virtual assistants, there are still a lot of issues when it comes to mixed language conversations, and the bot cannot comprehend the meaning or context of the conversation from the user end, and this results in user dissatisfaction, and human handover.

As any company, in order to make scalable progress in customer support, their main aim is to minimize human handover and perform successful conversations with user satisfaction using Conversational AI. In order to achieve this, the bot needs to understand the native human language, and how it can be mixed with other languages and yet be understandable.

Also, most virtual assistants these days are using only 3 human sentiments for holding conversations with the users. But the actual human sentiments extend way more than positive, negative and neutral. A user might be dissatisfied by the specific task performed by the bot, yet might want to work with, in the conversation, or might have a strong disagreement, etc. Human Emotions are complex, and we need a better classifier than just positive, negative and neutral to stand apart in the market.

For better understanding and improving the efficiency of the virtual assistant, it needs to constantly learn from the conversations held with the customer, and summarize the flow of the same. An efficient text summarizer is a must when it comes to learning from the new conversations and summarizing a long statement and identifying the intent from the same.

1.3 Objective(s)

With the progress so far in the bot development, we have classified our objectives into three parts, namely:

1. Efficient Text Summarizer
2. Sentiment Analysis
3. Mixed Language Conversations

This paper focuses on performance capabilities of the Text-to-text transfer transformer and different approaches for mixed code sentiment analysis and performs a comparative analysis on the various implemented models under the same.

2. Related Work (with Comparisons)

Our approach to build a single context conversational AI virtual assistant requires the use of Transfer Learning, where we apply an already pretrained model on general tasks, and train it on our specific tasks. This approach allows us to take use of State of the Art models, as well as no rigorous training on large amounts of datasets.

We explored a word vector approach for intent classification, but since the current transformer models have proved to show a much higher accuracy, we discarded the word vector approach.

Earlier Transformer based models were used for Machine Translation, but now they can be extended to any NLP text-to-text task. The primary building block of the Transformer is ***self-attention.*** Self-attention is a variant of attention that processes a sequence by replacing each element by a weighted average of the rest of the sequence.

It has recently also become common to use models consisting of a single Transformer layer stack, with varying forms of self-attention used to produce architectures appropriate for language Modeling or classification and span prediction tasks.

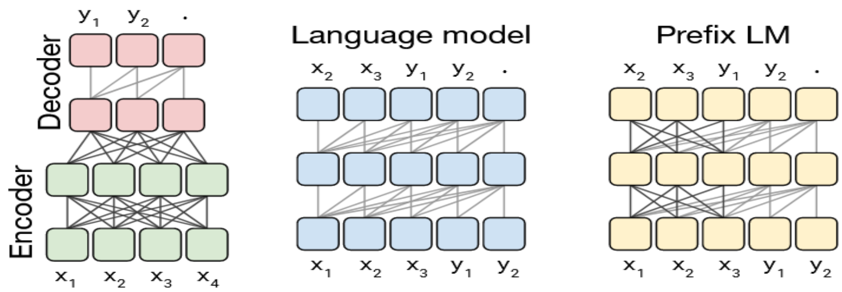


Figure 4: Schematics of the Transformer architecture variants we consider. In this diagram, blocks represent elements of a sequence and lines represent attention visibility. Different colored groups of blocks indicate different Transformer layer stacks. Dark grey lines correspond to fully-visible masking and light grey lines correspond to causal masking. We use "." to denote a special end-of-sequence token that represents the end of a prediction. The input and output sequences are represented as x and y respectively. Left: A standard encoder-decoder architecture uses fully-visible masking in the encoder and the encoder-decoder attention, with causal masking in the decoder. Middle: A language model consists of a single Transformer layer stack and is fed the concatenation of the input and target, using a causal mask throughout. Right: Adding a prefix to a language model corresponds to allowing fully-visible masking over the input.

While the Transformer was originally introduced with an encoder-decoder architecture, much modern work on transfer learning for NLP uses alternative architectures.

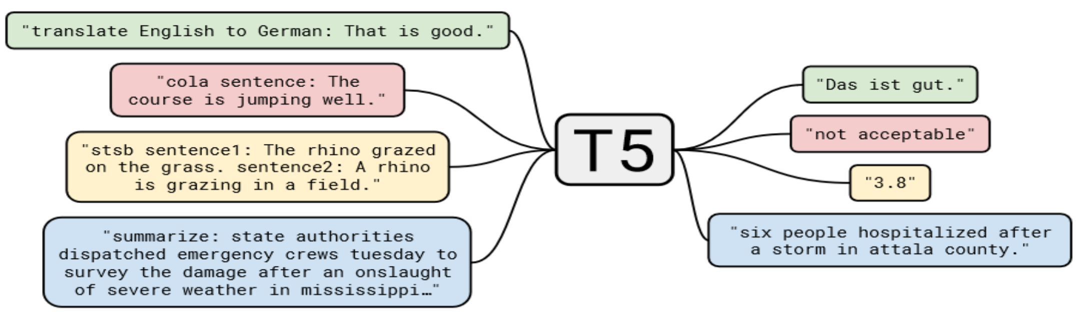


Figure 1: A diagram of our text-to-text framework. Every task we consider—including translation, question answering, and classification—is cast as feeding our model text as input and training it to generate some target text. This allows us to use the same model, loss function, hyperparameters, etc. across our diverse set of tasks. It also provides a standard testbed for the methods included in our empirical survey. "T5" refers to our model, which we dub the "**T**ext-**t**o-**T**ext **T**ransfer **T**ransformer".

Text-to-text Transfer Transformer can help achieve a wide variety of tasks from translation to classification, and it works by providing a text input to generate some target output.

After T5, our main focus diverted towards finding the optimum model for mixed code sentiment analysis. We explored Hierarchical Context Recognition Approach, trained transformer and ensemble models, details of which will be explored in the paper.

For the purpose of sentiment analysis in mixed code datasets, we had to perform many preprocessing steps, from keyword extraction, lemmatization, removal of stopwords, using translation API in each of the models for it to perform efficiently.

Hierarchical Context Modelling System, from its name suggests a hierarchical analysis of the input sentences. This approach uses CNN layers for obtaining the local context from the inputs and self attention layers at the end to obtain the global contexts. This model achieved the validation accuracy of 67% on sentimix dataset.

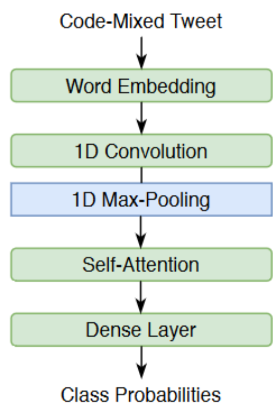


Figure 1: Model Architecture

Architecture of HCMS Model: https://arxiv.org/pdf/2007.12076

An ensemble model approach uses Naive Bayes, Multinomial Naive Bayes, Binomial Naive Bayes, Logistic Regression, Stochastic Gradient Descent, SklearnClassifier, Maximum Entropy, and gives an ensemble prediction. The overall accuracy of this approach came to be 62% for Hinglish Tweets scraped from the internet.

Another approach used for mixed code sentiment analysis was using a transformer model for sentiment analysis. We used XLM RoBERTa Base Model for the purpose, which uses adversarial layers, preventing overfitting of the datasets. This model achieved the validation accuracy of 60% upon training, proving that the Hierarchical Context Modelling System (HCMS) is the finest way for mixed code sentiment analysis.

3. Proposed Technique(s) and Algorithm(s)

We have created a pipeline for the purpose of text summarization and sentiment analysis from pretrained BERT Models obtained from Hugging Face Library.

We followed the recommended training instructions of hyperparameters to avoid any absurd results. BERTClassifier with Adam Optimizer was used for running the downstream sentiment analysis tasks, and for the purpose of training, tweets were extracted from the internet.

Hierarchical Context Modelling System (HCMS) uses a custom pipeline of Convolutional Neural Network Layers, followed by Max Pooling and Self Attention. The self attention layer uses sigmoid function as the activation, and the model uses Adam optimizer for training purposes.

```
model = Sequential()
model.add(Embedding(max_features+1, embedding_size, input_length=maxlen))
model.add(Conv1D(filters,
                 kernel_size,
                 padding='valid',
                 activation='relu',
                 strides=1))
model.add(MaxPooling1D(pool_size=pool_size))
model.add(SeqSelfAttention(attention_activation='sigmoid'))
model.add(Flatten())
model.add(Dropout(0.2))
model.add(Dense(3, activation='softmax'))
model.compile(loss='categorical_crossentropy',
              optimizer='adam',
              metrics=['accuracy'])
```

Model Architecture of HCMS

HCMS uses a preprocessing function like de-emojification, that replaces the emojis with the corresponding keywords, and replaces the acronyms to their full forms for the model to better understand the sentences.

The ensemble model for mixed code sentiment analysis uses Naive Bayes, Multinomial Naive Bayes, Binomial Naive Bayes, Logistic Regression, Stochastic Gradient Descent, SklearnClassifier and Maximum Entropy to predict the final sentiment. The downside of this model is that it performs only binary classification, although the preprocessing steps used here can be used as the motivation for various other models for sentiment analysis.

The preprocessing steps included cleaning the raw datasets, including de emojification, Lemmatization, Spellcheck, Identifying Stopwords, Feature Extraction: Adjectives, Nouns, etc, Google Translation API before finally being fed to the ensemble model for training and testing purposes. The training of the ensemble model is done on sentiment140 dataset from kaggle.

XLM Roberta is another mixed code sentiment analyzing model which uses a pre-trained transformer and performs fine tuning for sentimix dataset. The recommended hyperparameters were not disturbed which might affect the accuracy of the model. The model was trained for 5 epochs and the final accuracy came to be 61%, and the training was done with Adam optimiser.

- Batch size: 16, 32
- Learning rate (Adam): 5e-5, 3e-5, 2e-5
- Number of epochs: 2, 3, 4

```
The XLMRoberta model has 201 different named parameters.

==== Embedding Layer ====

roberta.embeddings.word_embeddings.weight                (250002, 768)
roberta.embeddings.position_embeddings.weight               (514, 768)
roberta.embeddings.token_type_embeddings.weight               (1, 768)
roberta.embeddings.LayerNorm.weight                             (768,)
roberta.embeddings.LayerNorm.bias                               (768,)

==== First Transformer ====

roberta.encoder.layer.0.attention.self.query.weight         (768, 768)
roberta.encoder.layer.0.attention.self.query.bias               (768,)
roberta.encoder.layer.0.attention.self.key.weight           (768, 768)
roberta.encoder.layer.0.attention.self.key.bias                 (768,)
roberta.encoder.layer.0.attention.self.value.weight         (768, 768)
roberta.encoder.layer.0.attention.self.value.bias               (768,)
roberta.encoder.layer.0.attention.output.dense.weight       (768, 768)
roberta.encoder.layer.0.attention.output.dense.bias             (768,)
roberta.encoder.layer.0.attention.output.LayerNorm.weight         (768,)
roberta.encoder.layer.0.attention.output.LayerNorm.bias         (768,)
roberta.encoder.layer.0.intermediate.dense.weight          (3072, 768)
roberta.encoder.layer.0.intermediate.dense.bias                (3072,)
roberta.encoder.layer.0.output.dense.weight                (768, 3072)
roberta.encoder.layer.0.output.dense.bias                       (768,)
roberta.encoder.layer.0.output.LayerNorm.weight                 (768,)
roberta.encoder.layer.0.output.LayerNorm.bias                   (768,)

==== Output Layer ====

classifier.dense.weight                                     (768, 768)
classifier.dense.bias                                           (768,)
classifier.out_proj.weight                                    (3, 768)
classifier.out_proj.bias                                          (3,)
```

Model architecture of XLM RoBERTa

4. Dataset(s) Used

For the purpose of Sentiment Analysis, we had a wide variety of datasets to choose from, but the aim was to find the datasets which could be used for most of the models for comparative analysis. Sentiment140 was one such dataset from Kaggle. It contains 1,600,000 tweets extracted using the twitter api, but the classification is binary in nature. For the training of the ensemble model, 10,000 examples from this dataset were extracted and an accuracy of 61% percent was obtained.

Sentimix Dataset is another open source Hinglish Dataset which was used for the training of the transformer model, and the classification was done in three classes, namely positive, negative and neutral. For the training of the HCMS model, tweets were scraped with the use of twitter API, and a small dataset was generated with sentiment labeling.

test_data

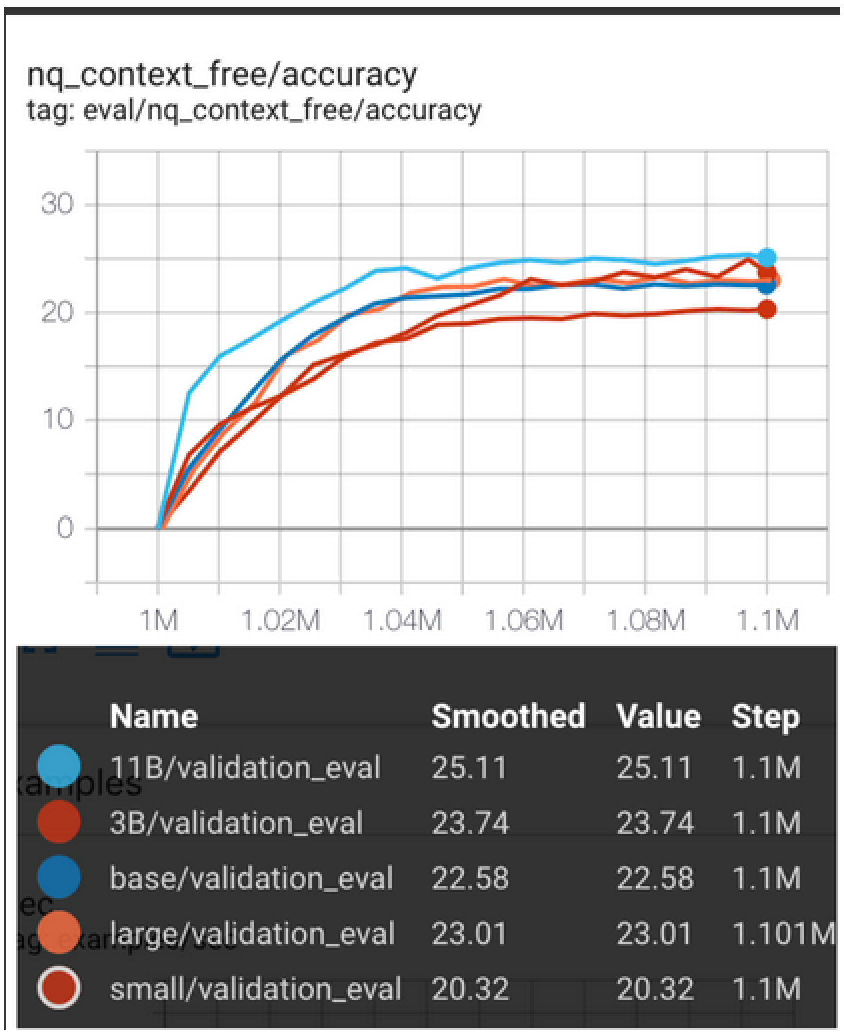
Label	number	date	no_query	name	Tweet
0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, that's a bummer. You shoulda got David Carr of Third Day to do it. ;D
0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by texting it... and might cry as a result School today also. Blah!
0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Managed to save 50% The rest go out of bounds
0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	Karoli	@nationwideclass no, it's not behaving at all. i'm mad. why am i here? because I can't see you all over there.
0	1467811372	Mon Apr 06 22:20:00 PDT 2009	NO_QUERY	joy_wolf	@Kwesidei not the whole crew
0	1467811592	Mon Apr 06 22:20:03 PDT 2009	NO_QUERY	mybirch	Need a hug
0	1467811594	Mon Apr 06 22:20:03 PDT 2009	NO_QUERY	coZZ	@LOLTrish hey long time no see! Yes.. Rains a bit ,only a bit LOL , I'm fine thanks , how's you ?
0	1467811795	Mon Apr 06 22:20:05 PDT 2009	NO_QUERY	2Hood4Hollywood	@Tatiana_K nope they didn't have it
0	1467812025	Mon Apr 06 22:20:09 PDT 2009	NO_QUERY	mimismo	@twittera que me muera ?
0	1467812416	Mon Apr 06 22:20:16 PDT 2009	NO_QUERY	erinx3leannexo	spring break in plain city... it's snowing
0	1467812579	Mon Apr 06 22:20:17 PDT 2009	NO_QUERY	pardonlauren	I just re-pierced my ears
0	1467812723	Mon Apr 06 22:20:19 PDT 2009	NO_QUERY	TLeC	@caregiving I couldn't bear to watch it. And I thought the UA loss was embarrassing . . . . .
0	1467812771	Mon Apr 06 22:20:19 PDT 2009	NO_QUERY	robrobbierobert	@octolinz16 It it counts, idk why I did either. you never talk to me anymore
0	1467812784	Mon Apr 06 22:20:20 PDT 2009	NO_QUERY	bayofwolves	@smarrison i would've been the first, but i didn't have a gun. not really though, zac snyder's just a doucheclown.
0	1467812799	Mon Apr 06 22:20:20 PDT 2009	NO_QUERY	HairByJess	@iamjazzyfizzle I wish I got to watch it with you!! I miss you and @iamlilnicki how was the premiere?!
0	1467812964	Mon Apr 06 22:20:22 PDT 2009	NO_QUERY	lovesongwriter	Hollis' death scene will hurt me severely to watch on film wry is directors cut not out now?
0	1467813137	Mon Apr 06 22:20:25 PDT 2009	NO_QUERY	armotley	about to file taxes
0	1467813579	Mon Apr 06 22:20:31 PDT 2009	NO_QUERY	starkissed	@LettyA ahh ive always wanted to see rent love the soundtrack!!
0	1467813782	Mon Apr 06 22:20:34 PDT 2009	NO_QUERY	gi_gi_bee	@FakerPattyPattz Oh dear. Were you drinking out of the forgotten table drinks?
0	1467813985	Mon Apr 06 22:20:37 PDT 2009	NO_QUERY	quanvu	@alydesigns i was out most of the day so didn't get much done
0	1467813992	Mon Apr 06 22:20:38 PDT 2009	NO_QUERY	swinspeedx	one of my friend called me, and asked to meet with her at Mid Valley today...but i've no time *sigh*
0	1467814119	Mon Apr 06 22:20:40 PDT 2009	NO_QUERY	cooliodoc	@angry_barista I baked you a cake but I ated it
0	1467814180	Mon Apr 06 22:20:40 PDT 2009	NO_QUERY	viJILLante	this week is not going as i had hoped
0	1467814192	Mon Apr 06 22:20:41 PDT 2009	NO_QUERY	Ljelli3166	blagh class at 8 tomorrow
0	1467814438	Mon Apr 06 22:20:44 PDT 2009	NO_QUERY	ChicagoCubbie	I hate when I have to call and wake people up
0	1467814783	Mon Apr 06 22:20:50 PDT 2009	NO_QUERY	KatieAngell	Just going to cry myself to sleep after watching Marley and Me.
0	1467814883	Mon Apr 06 22:20:52 PDT 2009	NO_QUERY	gagoo	im sad now Miss.Lilly
0	1467815199	Mon Apr 06 22:20:56 PDT 2009	NO_QUERY	abel209	ooooh.... LOL that leslie.... and ok I won't do it again so leslie won't get mad again
0	1467815753	Mon Apr 06 22:21:04 PDT 2009	NO_QUERY	BaptisteTheFool	Meh... Almost Lover is the exception... this track gets me depressed every time.
0	1467815923	Mon Apr 06 22:21:07 PDT 2009	NO_QUERY	fatkat309	some1 hacked my account on aim now i have to make a new one
0	1467815924	Mon Apr 06 22:21:07 PDT 2009	NO_QUERY	EmCDL	@alielayus I want to go to promote GEAR AND GROOVE but unfornately no ride there I may b going to the one in Anaheim in May though
0	1467815988	Mon Apr 06 22:21:09 PDT 2009	NO_QUERY	merisssa	thought sleeping in was an option tomorrow but realizing that it now is not. evaluations in the morning and work in the afternoon!
0	1467816149	Mon Apr 06 22:21:11 PDT 2009	NO_QUERY	Pbearfox	@julieebaby awe i love you too!!!! 1 am here i miss you
0	1467816665	Mon Apr 06 22:21:21 PDT 2009	NO_QUERY	jsoo	@HumpNinja I cry my asian eyes to sleep at night

Sentimix Dataset used for Transformer and HCMS Model

5. Experiment(s) and Result(s)

5.1 Evaluation and Result(s) (Comparisons)

We worked on Text-to-text Transfer Transformer (T5) as an alternative to BERT, and analyzed context free text to text generation using a Transformer. T5 is a text-to-text model, which enables us to train it on arbitrary tasks involving a textual input and output. As we showed in our paper, a huge variety of NLP tasks can be cast in this format, including translation, summarization, and even classification and regression tasks. More simplistic, yet same computational cost as compared to task specific architectures. Natural Questions is a challenging corpus for open-domain QA. Each example includes a question along with an entire Wikipedia article that may or may not contain its answer. The goal is to produce the correct answer given this context. In our case, we ignored the provided context in hopes that the model will learn to find the answers from the world knowledge it has acquired during pre-training.



These are the prediction accuracies obtained for various T5 models with different parameters. As we can see, these were State of the Art but quite low to be included and implemented within the virtual assistant. Although the error analysis approach here is verified for exact keywords and phrases within the predictions, a similar response might have been rejected due to the same reason.

```
Predictions using checkpoint 1025000:

Q: Where is the Google headquarters located?
A: in the silicon valley in the united states


Q: What is the most populous country in the world?
A: china


Q: Who are the 4 members of The Beatles?
A: john lennon, paul mccartney, john lennon jr. and george


Q: How many teeth do humans have?
A: thirty
```

These are the predictions the T5 Model made on the test questions from Natural Questions Dataset. All the training was done from a context free approach and only the existing knowledge of pretrained T5 Models was considered for testing.

The mixed code sentiment analysis model HCMS, received a final accuracy of 67% with the f1 score of 68% on Hinglish Dataset.

model	precision	recall	accuracy	f1-score
preds	0.681706	0.680411	0.676333	0.680694

The biggest upside of this model is that it does not require any pre-training and can be trained on proposed datasets directly, providing a high degree of accuracy and easy incorporation in the pipeline.

The XLM RoBERTa transformer model received a validation accuracy of 61% on the sentimix dataset, upon a training over 5 epochs.

It was noticed that upon training, the dataset began to overfit after the 3rd epoch and the validation loss increased linearly with each epoch. So, the net accuracy of 2nd epoch was considered as the true accuracy of the model.

The ensemble model which uses Naive Bayes, Multinomial Naive Bayes, Binomial Naive Bayes, Logistic Regression, Stochastic Gradient Descent, SklearnClassifier, Maximum Entropy gives a final accuracy of 62% with f1 score of 63%.

```
[[61.53333333333333, 'NB'],
 [61.4, 'MNB'],
 [62.4, 'BNB'],
 [63.266666666666667, 'LogReg'],
 [63.0, 'SGD'],
 [62.06666666666667, 'SVC'],
 [62.06666666666667, 'MaxEnt'],
 [62.666666666666664, 'Hybrid']]
```

```
Accuracy of hybrid :  62.666666666666664
Hybrid

              precision    recall  f1-score   support

    positive       0.67      0.48      0.56       738
    negative       0.60      0.77      0.68       762

    accuracy                           0.63      1500
   macro avg       0.64      0.62      0.62      1500
weighted avg       0.64      0.63      0.62      1500
```

The final accuracy of the ensemble obtained after training the model of sentiment140 dataset/

```
[ ]   1 func("arrey waah! I'm very proud of you")

['proud']
Tweet given by user :  Arrey Waah!I'm Very Proud of You
naive bayes classifier
This Tweet is  positive
------------------------------
Multinomail naive bayes classifier
This Tweet is  positive
------------------------------
Bernouli classifier
This Tweet is  positive
------------------------------
Bernouli LogisticRegression_classifier
This Tweet is  positive
------------------------------
SGD classifier
This Tweet is  positive
------------------------------
SVC classifier
This Tweet is  positive
------------------------------
Max Entropy classifier
This Tweet is  positive
------------------------------
Hybrid model
This Tweet is  positive
```

The hybrid model correctly predicting the sentiment of mixed code sentence

5.2 Error Analysis

For the case of the T5 Model, the measure of error analysis was flawed and we couldn't obtain accurate results. The accuracy was measured against exact phrases or keyword predictions, and didn't take into account the synonyms or phrases with similar meanings. Due to the same reason, the accuracy came out to be as low as 25%.

Instead if a more liberal error analysis approach could have been considered where the final prediction was matched against the actual value and based on the similarity index, the accuracy was detected.

Mixed Code sentiment Analysis accuracy could be measured quite easily since it's a classification task, although it becomes difficult to compare the results of different models only on the basis of their accuracies. The ensemble model performs quite well even after training on just 10,000 training examples, but since it performs only binary classification, on the other hand, the pre-trained model of XLM RoBERTa gives an accuracy of 61% on sentimix dataset, with positive, neutral and negative sentiments. The HCMS model, though having a comparatively simpler architecture, performs better than the other models with approximately 68% accuracy.

6. Conclusion and Future Work

We have explored the State of the Art model architectures and integrated them keeping in mind that the pipeline remains time efficient, and can perform live conversations with the user without considerable delay. We have currently integrated text summarizer and sentiment analyzer to the previously built conversational AI bot, and we are working on extending the current model to mixed code or mixed language conversations.

The mixed code sentiment analysis task has achieved a maximum accuracy of 68% with Hierarchical Context Modeling System, followed by 62% accuracy on binary classification of mixed code sentiment analysis, and 61% accuracy by XLM RoBERTa model.

The accuracy could be further improved by addition of more training datasets, but due to limited availability of mixed code datasets, this couldn't be done this semester.

The following future work could be performed:

- Due to its diverse applications, T5 Model could be used for a wide variety of tasks from sentiment analysis to text summarization and fixed response generation.
- More transformer models could be explored for the purpose of mixed code training for sentiment analysis and text summarization.
- Accuracy of current models could be increased with further training on mixed code datasets.
- If the Hindi words are converted to Devanagari Script from the Roman script, the redundancy could be avoided and accuracy could be further increased. Identification and conversion of scripts could be performed.

7. Work Update

7.1 Work completed

We have constructed a pipeline for text summarization and sentiment analysis on a single context conversation and are able to perform them with high accuracy on a single language based conversation. For the purpose of mixed code sentiment analysis, we have implemented three different approaches and achieved the highest accuracy of 68%. We also explored the uses of the T5 model and it;s capabilities could be used for a wide variety of tasks from sentiment analysis to response generation.

7.2 Work Remaining (Future Work)

We have summarized the future objectives in the Question Answering system

- Due to its diverse applications, T5 Model could be used for a wide variety of tasks from sentiment analysis to text summarization and fixed response generation.
- More transformer models could be explored for the purpose of mixed code training for sentiment analysis and text summarization.
- Accuracy of current models could be increased with further training on mixed code datasets.
- If the Hindi words are converted to Devanagari Script from the Roman script, the redundancy could be avoided and accuracy could be further increased. Identification and conversion of scripts could be performed.
- Aggregating Different Mixed Code datasets into a single dataset to train the existing models and get a better comparison and inclusion of more preprocessing steps like lemmatization to further increase the accuracy.

8. References

- https://www.ibm.com/cloud/learn/conversational-ai
- https://arxiv.org/pdf/1910.10683v3.pdf
- https://github.com/google-research/text-to-text-transfer-transformer
- https://github.com/google-research/text-to-text-transfer-transformer/blob/main/notebooks/t5-deploy.ipynb
- https://github.com/google-research/text-to-text-transfer-transformer/blob/main/notebooks/t5-trivia.ipynb
- https://arxiv.org/pdf/1911.03437v5.pdf
- https://github.com/Rajan021/CodeMixing-Sentiment-Anaysis/blob/master/code_mix_final.csv
- https://colab.research.google.com/github/YarikMR/Multi-Language_Sentiment_Classifier/blob/master/MSE_post.ipynb?authuser=1#scrollTo=W6ohHPLKWvxi
- https://github.com/somnath-banerjee/Code-Mixed_SentimentAnalysis
- https://medium.com/swlh/gradient-based-adversarial-attacks-an-introduction-526238660dc9
- https://arxiv.org/pdf/2007.12076
- https://github.com/sakshidgoel/Bilingual-Sentiment-Analysis
- https://github.com/IamAdiSri/hcms-semeval20
- https://github.com/rsgoss/NLP_finalproj