Financial Complaint Summarization using Reinforcement Learning Techniques

Reinforcement Learning Project

School of Computer Science & Engineering (SCOPE)

Submitted By:

1. **Govindgari Sai Srushik (21BCE8659)**
2. **Medi Pavan Teja (21BCE9805)**

Submitted To:

Dr. Suma Kamalesh Gandhimathi

TABLE OF CONTENTS

Abstract

The plethora of client complaints in the financial business is a formidable obstacle for establishments aiming to improve customer contentment and adherence to regulations. The sheer amount and complexity of financial complaints necessitate the use of effective and efficient summarizing techniques in order to extract crucial data and insights. sometimes it might be a possibility that the complaint description is vast, pretending the respondent to skip the complaint and move to the next.

In this research we are here to propose a new hybrid extractive-abstractive model [1] for “financial complaint Summarization model” that adopts the capabilities of the reinforced Q - Learning which combines with BERT (Bidirectional Encoder Representations from Transformers) word embedding and T5- small Transformer. Which can be efficiently summarize the complaint description. Firstly, we have developed a extractive text summarization model using Bi-directional Encoder Representational Transformer (BERT) model and abstractive text summarization model using T5-Transformer model, then we optimized that hybrid model with the reinforced Q-Learning Algorithm. We made use of the FINancial Complaint CORpus (FINCROP) dataset and the ROUGE (Recall-Oriented Understudy for Gisting evaluation) score metric as an evaluation strategy for validating the model's performance.

Introduction

The technology of automatic text summarization has numerous applications across various sectors. Users often complain about their questions, technical glitches, etc., particularly within the financial sector. A massive amount of data will unavoidably be produced every day due to the rapid development and change of information techniques, particularly for mobile internet. These days, it is challenging to meet the demands of information retrieval in people's daily lives using the manual summarizing technique that was once common. As a result, with the flood of mass information, Automatic Text Summarization (ATS) is becoming more and more crucial.

There is now a significant surge in the amount of text, image, audio, and video content available due to the explosion of complaint information. To enable the respondents to swiftly comprehend and review the content, we must condense and summarize the vast amount of data and generalize the primary points that caught their interest.

The text summarization technique can be categorized into two main paradigms: extractive and abstractive. The extractive method involves extracting important sentences or sections of text from the original source and combining them to create a summary. [1,2] On the other hand, the abstractive method generates new words that may not exist in the source text while still preserving the original meaning.

In terms of differences, the extractive paradigm is relatively straightforward and ensures grammatical correctness, but it may lack consistency in terms of semantics. On the other hand, the abstraction paradigm is more concise but may include redundant information.

When summarizing a lengthy text, the extractive approach may be too simplistic and result in poor readability. Conversely, the abstractive method, which compresses a long input sequence into a fixed-length vector, may lead to information loss. Neither of these methods is particularly effective for summarizing lengthy texts.

2.1 Objective

Our main objective is to ease the work financial bodies to efficiently manage the complaints that they come across through the user's by making use of complaint description data. Which can be done by summarizing the complaint description of user. so, that the respondent can resolve the complaint. this can be achieved by the domain specific "Financial Complaint Summarization model."

Also, our objective is to use the hybrid extractive- abstractive model, which extracts the important sentences of the complaint description and on other-side abstractive method which preserves the meaning of the original sentence.

The pre-trained representation of BERT (Bidirectional Encoder Representations from Transformers), which is effectively implemented in many NLP tasks, is used as the text representation of all our models. This allows us to take advantage of the advantages of the pre-trained language model in the view of the forementioned issues. To produce better word embedding, BERT has been pre-trained on an extensive amount of unlabeled data. This BERT model acts as the extractive method of our hybrid model.

Furthermore, we can characterize the problem as an agent (the summarizer) interacting with the environment (the complaint text) to learn a policy that produces the precise summary of the complaint by replacing the current policy with the optimal policy in the context of complaint summarization using Reinforced Q-learning.

2.2 Problem Statement

Financial institutions receive a significant volume of complaints from customers regarding various issues such as transactions, account management, customer service, etc. Analyzing and summarizing these complaints efficiently is crucial for the institutions to address concerns and improve customer satisfaction.

The aim of this project is to develop a system that utilizes BERT, a Transformer algorithm & Q-learning, a reinforcement learning algorithm combinely, to automatically summarize financial complaints. The system should be able to process a large number of complaints, understand the context, and generate concise and informative summaries.

The major challenges that include while building this hybrid model is to address the potential biases in the training data, ensuring the summarization model is producing the consice and informative summaries, also handling the diversity of the financial complaints i.e., whether the complaint is negative, neutral, or positive.

The expected outcomes of the model is to efficiently understand and summarize the financial complaint, to enhance customer satisfaction through the effective complaint resolution.

2.3 Scope & Motivation

The reason behind this Research is prompted by the surge in complaints in the financial sector, which leads to consumer disappointment with customer service about a variety of issues brought on by technical difficulties, inquiries, etc. The respondent may take some time to evaluate the complaint description and tweak their response due to the lengthy description provided by the consumer. Therefore, effectively summarizing them can aid in comprehending client worries and enhancing services. Financial complaint summaries might result in faster resolution times, which will enhance the clientele's experience in the end.

The project scope is determined and can be achievable by making use of the FINancial Complaint CORpus (FINCROP) dataset[1]. A well defined which covers a wide range of complaints from Twitter platform through Twitter Scraping. Data wrangling over the dataset is the first step that has to be performed in order to obtain the good summary of the complaint description. here in this dataset removing stop-words, handling the irrelevant information like hashtags, mentions etc.

The Integration of BERT per-trained model helps in understanding the context of the complaints. while the Q-Learning can be applied to optimize the summarization process, by defining the reward function and policy over in which agent(summarizer) is interacting with the environment (complaint text).

Furthermore, as we can also find the images, videos along with the complaint description (text data). so, we can extend this model to a multimodal system, which efficiently deals with the image, audio, & video data of by combining with text data. these multi-modal systems have capability of understanding the context of the complaint more effectively.

Literature Survey

Loukia Avramelou et al. (2023) [3] conducted research involving automated text summarization to simplify the massive volume of textual material without abandoning information. Modern financial text summarization models claim that the Transformer can transform and evaluate text data effectively. They suggested an automated process for the improved text summarizing models. They noticed that by fine-tuning these LLMs, one may adapt to new domains and enhance oneself. It was also determined that using reinforced approaches to fine-tune Transformer based summarization models is a fascinating approach, however it is difficult to evaluate the introduced pipeline in other domains.

An insightful study on abstractive text summarization integrating reinforcement learning and attention mechanisms was conducted by Yash Kumar Atri et al. (2023) [4]. This work led to the development of the REISA model, which makes use of policy gradients and reinforced selective attention span. The key elements of the output are drawn out by this attention score. They came to the conclusion that the model significantly surpassed the baseline summary in terms of ROUGE and BERTScore.

A study was conducted on optimizing the multi-document summarization by DiJia Su et al. (2023).[5] This emphasizes the text redundancy. Sentence extraction is accomplished by identifying the key informational chunks and removing repetitions. Ploicy Blending maximum marginal relevance and Reinforcement Learning (PoBRL) can be used to achieve this. Their exceptional performance in multi-document summarization has been demonstrated by their empirical examination on PoBRL.

Additionally, an another optimization method has been provided to maximize Attention in sequence modeling. Deep reinforcement learning is used to automatically optimize attention distribution in order to reduce training losses, as shown by Hao Fei, Yue Zhang, et al. [6]. In order to give greater attention to words that are more informative, the agent is built to modify the attention weights by interacting with the data. Finally, had a conclusion that their RL model is inefficient while training.

HEEWON JANG et al. (2021) [7] defined a novel kind of reward function for a reinforcement learning algorithm and demonstrated a study on the quality of a summary of sentences. As altered versions of ROUGE-L, they introduced ROUGE-SIM & ROUGE WMD. The ROUGE-L function is not as effective as these functions. Additionally, the model's validation of reward functions is restricted to document information from other domains and it employs a single-layered LSTM, which has fewer trainable parameters.

An another extractive-abstractive model had introduced by WENFENG LIU et al. (2021) [2] to aim the limitation of document level summarization of sentences, which was done by sentence similarity matrix to extract important sentences out of document. Finally, they solved the poor readability of the generated summarization. Also stated that their model is not suitable for multi-document & cross-document summarization.

Ryosuke Kohita et al. (2020) [8] presented a novel method based on Q-learning with an edit-based summarizing in the field of abstractive text summarization. This technique creates an Editorial Agent and Language Model converter (EALM) by combining two essential modules. Edit actions such as replace, keep, and delete are predicted by the agent, and based on the action signals, the Language Model converter deterministically creates a summary. Using 30,000 randomly selected data points from the Gigaword corpus, the researchers discovered that EALM outperformed the earlier encoder-decoder-based techniques. Through qualitative analysis, they discovered that each unsupervised model has unique limits and that the quality of the summaries it generates is insufficient.

Neural sequence learning techniques have advanced significantly in single document summarizing (SDS), however Yuning Mao et al. (2020) [9] discovered that these techniques yield poor results in multi-document summarization (MDS). The reason for this is that MDS involves a bigger search space. To address this, they introduced Maximal Margin Relevance-guided Reinforcement Learning for MDS (RLMMR), which combines statistical metrics and sophisticated neural SDS techniques. By narrowing the searchspace and focusing MMR advice on a smaller number of highly promising candidates, RLMMR improves representation learning. On benchmark MDS datasets, extensive studies show that RL-MMR achieves state-of-the-art performance.

Yaser Keneshloo et al. (2020) [10] provided a general overview of a specific type of deep learning models called sequence-to-sequence(seq2seq) models and discussed some of the recent advances in combining training of these models with RL techniques. In their paper, they summarized some of the most important works that tried to combine these two different techniques and provided an open-source library for the problem of abstractive text summarization that shows how one could train a seq2seq model using different RL techniques.

An encoder-decoder model based on a double attention pointer network (DAPT) was proposed by ZHIXIN LI et al. (2020) [11]. In DAPT, the pointer network and soft attention provide more coherent core material, the self-attention mechanism gathers important information from the encoder, and the combination of both produces accurate and coherent summaries. Additionally, the problem of duplication is tackled and the quality of the generated summaries is enhanced by the utilization of the enhanced coverage technique. The suggested model was tested with results on the CNN/Daily Mail and LCSTS datasets. In comparison to the conventional pointer-generator model, the experimental findings demonstrate that the method can produce a more accurate and consistent summary and has enhanced the ROUGE evaluation index.

In order to integrate BERT (Bidirectional Encoder Representations from Transformers) word embedding with reinforcement learning for text summarization, Qicai Wang et al. (2019) [1] presented a novel hybrid extractive-abstractive model. Using the CNN/Daily Mail dataset, we compared the model's performance to that of the widely used automatic text summary model at the moment. The evaluation approach we employed was ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metrics. and discovered that the suggested model produced the greatest outcomes.

A novel extractive method for document summarizing based on a Deep Q-Network (DQN) was presented by Kaichun Yao et al. (2018) [12]. DQN can train a policy that maximizes the Rouge score with regard to gold summaries and represent the salience and redundancy of phrases in the Q-value approximation. Their tests demonstrate that, even in the absence of language annotation, the suggested method outperforms or is on match with state-of-the-art models on these datasets. We intend to investigate a word-level-based extraction technique using deep reinforcement learning in the future.

In response to the inconsistencies arising from flawed evaluation metrics, Scott M. Jordan et al. (2020) [16] introduced an innovative evaluation methodology tailored for reinforcement learning (RL) algorithms. Their methodology is designed to yield dependable performance measurements both within individual environments and when aggregated across a spectrum of environments. The authors validated their approach using their own dataset, subjecting a diverse array of RL algorithms to evaluation on standard benchmark tasks. Moreover, they conducted a comparative analysis of eleven algorithms, executing each one across various environments for 10,000 trials. Through succinct summaries and a variety of graphical representations, they presented the performance profiles of all algorithms. This methodological proposal offers a robust framework for facilitating future research by enabling more effective comparisons of RL algorithms.

Existing Methods and Disadvantages:

In the view of "Text Summarization", there are many various techniques have been developed and implemented to combine extensive amounts of textual data into succinct overviews. The approaches taken by these strategies range, from complex algorithmic models driven by natural language processing (NLP) to manual summary techniques.

This section gives a summary of the main current techniques for text summarization and talks about each one's drawbacks. Through an analysis of these techniques and their limitations, we hope to draw attention to the difficulties and areas that still need work in the field of automatic text summarization.

1. Manual Summarization:

- Traditional method involving human effort to condense lengthy text into shorter summaries.

Disadvantages:

1. **Time-consuming:** Requires considerable time and effort, especially for large volumes of text.
2. **Subjectivity:** Summaries may vary significantly depending on the individual's interpretation and bias.
3. **Scalability:** Not feasible for processing vast amounts of data efficiently.

2. Extractive Summarization:

- An algorithmic method that takes significant sentences or phrases from the source text and shows them based on Metric used.

Disadvantages:

1. **Lack of Cohesion:** Extracted sentences may not always flow well together, leading to disjointed summaries.
2. **Redundancy:** Often includes redundant information, as it directly extracts from the source text without rephrasing.
3. **Difficulty Handling Complex Texts:** Struggles with summarizing texts containing nuanced or ambiguous language.

3. Abstractive Summarization:

- Utilizes natural language processing (NLP) techniques to generate summaries that may contain rephrased and synthesized content not present in the original text.

Disadvantages:

1. Quality Concerns: Generates summaries that may contain grammatical errors, factual inaccuracies, or lose the essence of the original text.

2. Resource-Intensive: Requires significant computational resources and sophisticated algorithms, limiting its accessibility.

3. Training Data Dependency: Performance heavily relies on the availability of large and diverse training datasets, which might not always be readily accessible or suitable for all domains.

4. Hybrid Method of Summarization:

- Hybrid Method involves in combining both the extractive and abstractive methods to leverage their respective strengths.

Disadvantages:

1. Complexity: Integrating multiple techniques increases the complexity of the summarization process, requiring careful optimization and tuning.

2. Performance Trade-offs: Balancing between extraction and abstraction often involves trade-offs between summary quality and coherence.

3. Resource Demands: Hybrid models may require more computational resources and longer training times compared to single-method approaches.

Proposed Method & Advantages

Text Summarization, the process of extracting the key information while preserving it’s meaning and presenting it as a “Overview” of the entire large volume of text. Here in our work, our goal is to summarize the complaint description by customers and make it productive for respondent to solve the problem in a time bound.

In this research we aim to propose a novel approach for extracting the summaries of large text data through a hybrid model with Deep Reinforcement Learning techniques i.e., “Extractive - Abstractive Summarization” model with “Deep Q-Learning technique.”

This overcomes the problems of traditional methodologies of extracting summaries with human’s effort, and also the current disadvantages of extractive summarization models, like lack of cohesion, redundancy, handling complex texts, and also the disadvantages of abstractive summarization models, like quality, resource intensity, data dependency.

1. Dataset Collection:

We used the FINCROP dataset for our research, this data is openly available at git repository published by A. Singh, et.al [13]. The Financial Complaints Corpus (FINCROP), a compilation of annotations from complaints made in English on Twitter about financial organizations and customers. The relevant sentiment, emotion, and complaint severity classes have been added to the dataset. The dataset includes 3133 non-compliant cases and 3149 complaints from more than ten different domains (such as credit cards and mortgages).

This dataset has 6280 entries spread over 7 different features, (Domain, Complaint/ opinion, Complaint Label, Complaint_Cause, Severity level, Sentiment, Emotion). This dataset has complaint from 25 different domains with positive, neutral, and negative complaint label having 4 different levels of severity, 3 types of sentiments and with 7 types of emotion. This dataset is well balance and has a decent number of entries (i.e., 6280) which allows us to generate the summary of the complaint efficiently. To ensure unbaised condition we splitted the dataset into train and test set.

2. Data Preprocessing:

Data preprocessing is first and foremost task that we perform in order to train our model effectively. Data preprocessing involves Text Cleansing, Tokenization, stop word removal, word embeddings etc.

As the data is extracted from the twitter, in the complaint description there is a possibility that it may have mentions and hashtags that are considered as noisy data, which may resist our model from generating an effective summary of the complaint. we use regular expressions to remove those noise from text data.

After performing the text cleansing, we convert large complaint description into the tokenized words and remove the stop words from those tokenized words, to focus on more meaningful and informative words in the complaint. then these words are embedded into the vectors using Term Frequency – Inverse Document Frequency, to quantify the relevance of words in the complaint description.

3. Model Development:

We came up with a novel hybrid architecture for efficient text summarization that involves in utilizing the capabilities of Deep Reinforcement Learning Techniques i.e., Deep Q-Learning, which takes text as input and is passed to both the extractive and abstractive summarization models parallelly, and then gives the output accordingly, extractive summarizer extracts the important sentences or phrases from the text, while abstractive summarized generative the actual summary of the text.

These models are evaluated based on the ROUGE score of respective models. moreover, these ROUGE scores are passed onto the Q-Learning algorithm along with the summarized text. Where the Algorithm decides the whether to consider extractive summary or abstractive summary or the combination of both.

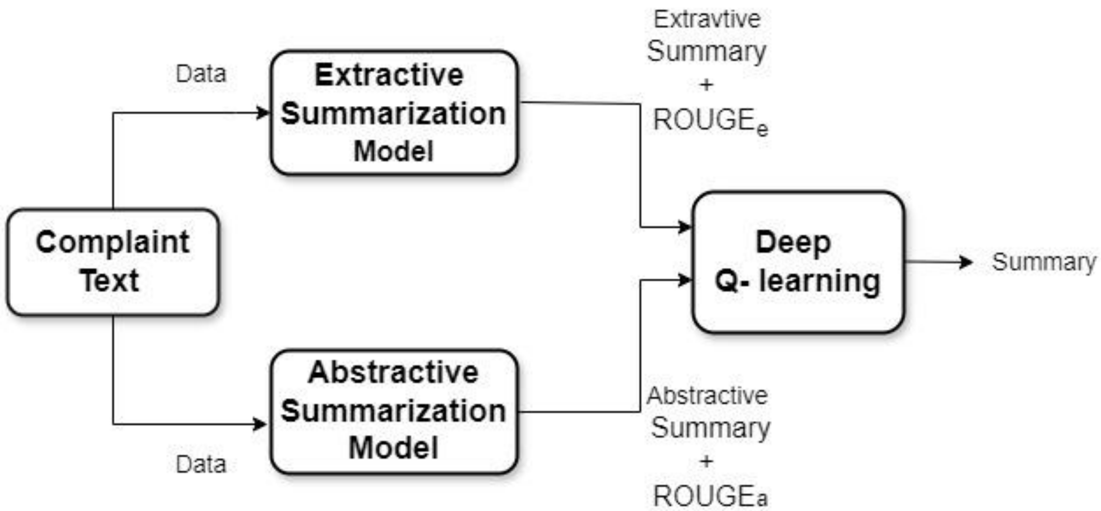

Figure 1: Depicts the proposed Training Methodology

1. Extractive Summarization:

Transformers are employed in sequence-to-sequence modeling to discover the long-range dependencies among words in a sentence. By addressing drawbacks such as variable-length input, parallelization, vanishing or inflating gradients, large data sets, etc., transformers outperformed other models in their field of application. It employs a neural architecture-based attention mechanism to dynamically highlight pertinent input data features, narrowing its focus to the essential elements or words.

In the view of extractive summarization, we utilized the pre-trained Bidirectional Encoder Representations from Transformers (BERT), while it was basically developed for question answering and sentiment analysis, it can also be adapted for text summarization purposes. As the two distinct semantic meanings will be distinguished and captured by the BERT embedding, which will generate two distinct vectors for the same word.

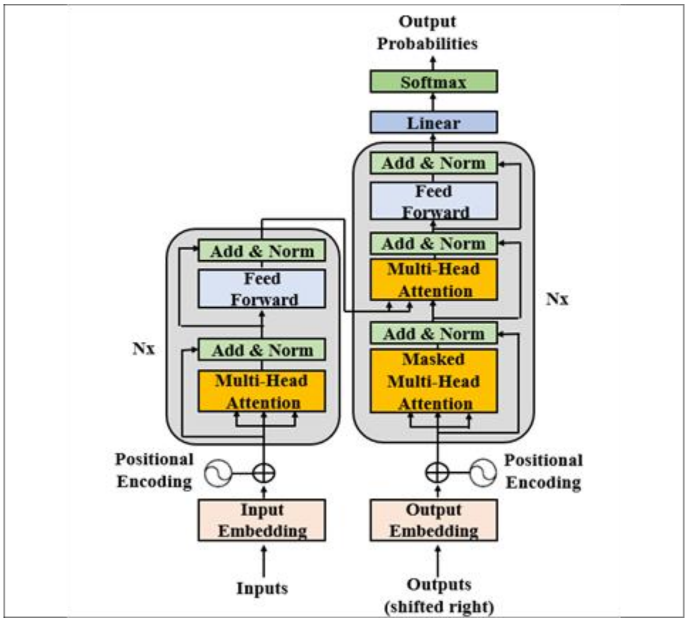

Figure 2: BERT model Architecture[14]

In order to enable the model to determine the relative relevance of each word in the input sequence while producing the summary, BERT incorporates self-attention layers into both the encoder and decoder blocks. This attention mechanism aids BERT in comprehending the text's context and capturing long-range dependencies.

2. Abstractive Summarization:

The abstractive summarization model is the Text-To-Text Transfer Transformer (T5) model architecture, which was presented by Colin Raffel et al. [15] and is intended for various language processing applications. This model architecture unifies several NLP activities under a common paradigm, making the training and evaluation process simpler. The FINCROP dataset is used to fine-tune the T5 model to perform summarization task.

To achieve the best possible performance for text summarization, hyperparameter variables such as batch size and learning rate are carefully adjusted. For effective computation and learning, a batch size of 8 is selected during the fine-tuning procedure. In order to strike a compromise between model optimization and convergence speed, a learning rate of 2e-5 is also used. This methodology ensures both quick learning and ongoing improvement throughout the training process.

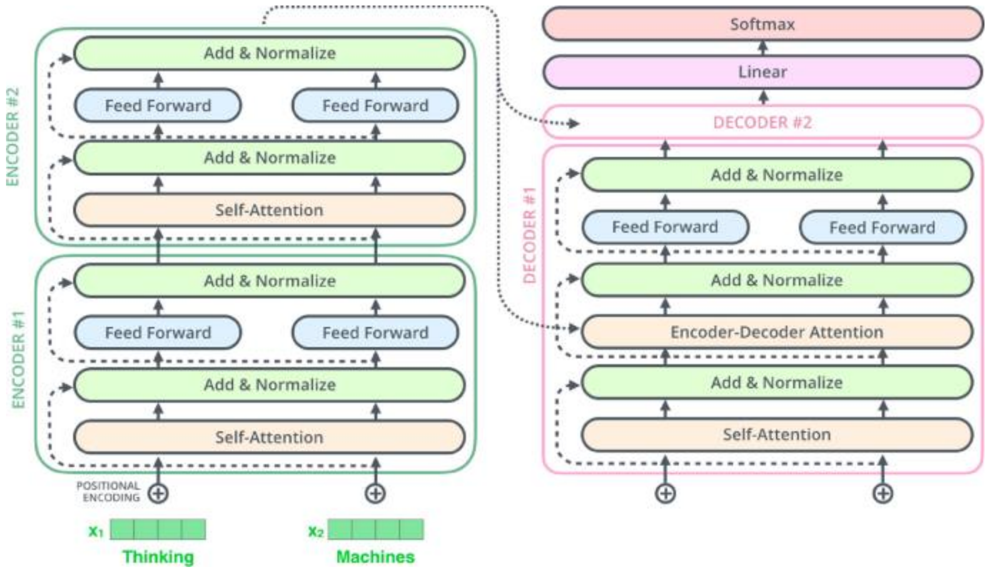

Figure 3: T5 Model Architecture [16]

3. Q-Learning Algorithm:

Q-learning is an off-policy, value-based, model-free algorithm that determines the optimal course of action given the agent's present condition. which train the value function to determine which state is more valuable and to take action, and which learn the consequences of their actions via experience without transition and reward function. The Q-Learning works well for dynamic environment, but we are pruned to use the Q-Learning Algorithm for a static environment i.e., for the finding the optimal text summary.

Further on, we use Deep Q-Learning to decide whether to select the extractive summary or the abstractive summary or the combination of the both the summaries based on the optimal policy (π*) of the agent to choose the optimal action, however there are only two optimal actions extractive Rouge score $ROUGE_{(e)}$ and abstractive Rouge score $ROUGE_{(a)}$.

Hyper-Parameters for Q-Learning:

1. State (**S**):

$$S = set\ of\ all\ original\ sentences\ or\ text$$

$$S_0 = \ input\ sentence\ or\ text$$

2. Action (**A**):

$$A = [ROUGE_{(e)},\ ROUGE_{(a)}]$$

3. Reward (**R**):

R_0 = null

$$R_{t+1} = \begin{cases} ROUGE_{(e)}, & if\ Se \\ ROUGE_{(a)}, & if\ Sa \end{cases}$$

In the context of text summarization, agent chooses optimal action based on the optimal policy. In our approach the optimal policy for choosing the optimal text summary is the value of maximum of the both the extractive and abstractive Q-Function values i.e., $Q_e\ (S_t, A_t)$ and $Q_a\ (S_t, A_t)$ that can maximize the expected rewards of both. where S_t and A_t are the state and action at time stamp "t'.

Optimal Policy,

$$\Pi^*(a) = \max\ \{Q_e\,(S_t, A_t), Q_a\,(S_t, A_t)\} \text{ ------ (1)}$$

Where the Q-Functions for Extractive and Abstractive models are,

Q-Function for extractive model,

$$\mathrm{Qe\ (St, At),} \leftarrow \mathrm{Qe\ (St, At),} + \alpha\ [\mathrm{Rt} + 1 + \gamma \max \mathrm{Qe\ (St} + 1, \mathrm{a}) - \mathrm{Qe\ (St, At)}\ ] \text{ -------(2)}$$

Q-Function for abstractive model,

$$\mathrm{Qa\ (St, At),} \leftarrow \mathrm{Qa\ (St, At),} + \alpha\ [\mathrm{Rt} + 1 + \gamma \max \mathrm{Qa\ (St} + 1, \mathrm{a}) - \mathrm{Qa\ (St, At)}\ ] \text{ -------(2)}$$

The optimal action can be selected is based on the $\pi^*(a)$

$$at = \begin{cases} ROUGEe, & if\ \pi * (a) = Qe(St, At) \\ ROUGEa, & if\ \pi * (a) = Qa(St, At) \end{cases}$$

And the optimal summary is,

$$St + 1 = \begin{cases} extracrtive\ summary, if\ at = ROUGEe \\ abstracrtive\ summary, if\ at = ROUGEa \end{cases}$$

The optimal summary obtained at state S_{t+1} is final optimal summary for the given input complaint text.

Results

In our research we observed that the evaluation of the extractive and abstractive summarization models achieved expected ROUGE scores, and the mean of scores were 90.60(ROUGE-1) and 84.40(ROUGE-1). However, the output summary given by the Q-Learning algorithm will achieve only a hypothetical decent score due to maximization which needs to be implemented. And, there are many metrics to be analyzed like Precession, f1-score to get the optimal policy, that maximizes the expected reward to get optimal summary.

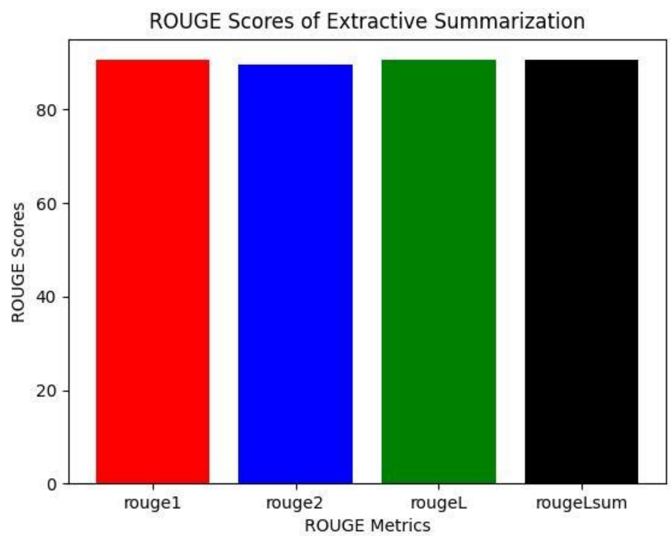

Figure 4:ROUGE Scores of Extractive Summarization

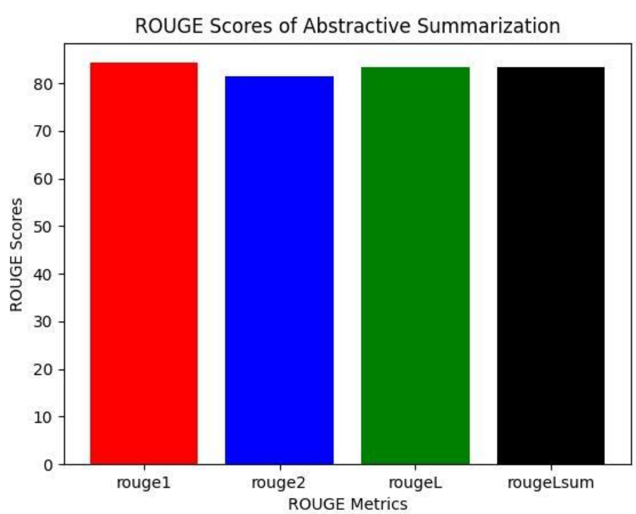

Figure 5: ROUGE Scores of Abstractive Summarization

Conclusion/ Future Works

There is a need of restructuring of optimal policy is required as it leads to the maximization problem in Q-Learning to maximize the rewards. Also, we know the basic problems of Q-Learning i.e., needs a large memory to store and update the Q-Table as the number of states increases and another, is that the amount of time required to explore new states. To overcome this problem “Deep Q-Networks” can be used with “Experience Replay” techniques. Moreover, modern techniques can be used to evaluate the performance of the reinforcement learning algorithms [17]. Apart from these we can combine this model with image, video and audio data making it as a “Multimodal System” as there is a large Complaint data available in those forms.

Project Flow/ Framework of the Proposed System

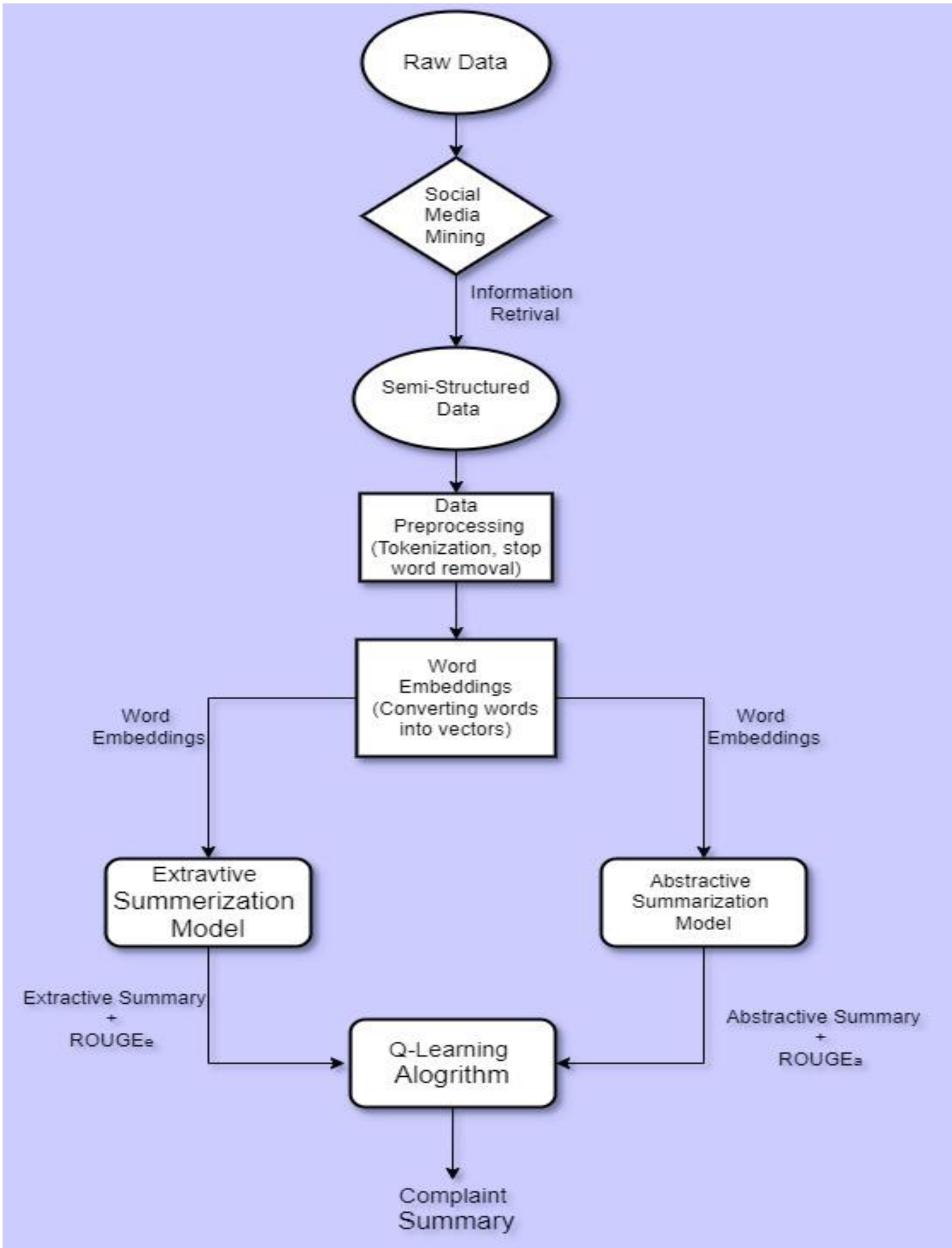

Figure 6: Project Framework Flow

1. Raw Data:

- The Complaint description given by customer's is extracted from the Twitter Platform.

2. Social Media Mining:

- To extract the data from the Twitter platform we perform Information Retrieval Techniques, to extract the necessary complaint data from the twitter posts across the domains. Through which we get the Semi-Structured Data.

3. Data Preprocessing:

- The Data obtained by performing Social Media Mining, has some noisy data like mentions and hashtags, which resist our model's performance. So, try to remove those noisy data.
- Later, we perform Tokenization, and remove Stop words, to focus more on key words.

4. Word Embeddings:

- These words are further converted into the vectors using word embedding techniques like TF-IDF, the vectors quantify the relevance of the words in the complaint description.

5. Extractive Summarization:

- These word embeddings are passed onto the extractive summarization model to get extractive summary.
- Extractive summarization model extracts the important sentences as the summary.

6. Abstractive Summarization:

- Also the word embeddings are passed onto the abstractive summarization model to generate abstractive summary.
- Abstractive summarization model generates the actual summary of the sentences.

7. Q-Learning Algorithm:

- The Q-Learning algorithm takes the input summaries of both the summarization models along with their respective ROUGE scores, based on which it determines the whether to consider extractive summary or abstractive summary or combination of both. Then returns the output summary.

Hardware and Software requirements

Hardware Requirements:

- **GPU:** As we are dealing with the pre-trained large language models (LLMs) which were trained on large text data it requires a more Computational Power. Here in this project, we are working on Google Collab Environment, where we can utilize the cloud enabled T4-GPU.

- **RAM:** This project requires a good amount memory, in order to process the data very quickly. This is prebuilt available as a Cloud RAM from Google Collab Environment.

Software Requirements:

- **Numpy & Pandas:** These are the Python Libraries which can be used for Data Manipulation.

- **Matplotlib & Seaborn:** These are the Python Visualization Libraries, that can be used to visualize the data.

- **NLTK:** NLTK is a popular Natural Language Processing library, which can be used to make operation over the text data in order to make it suitable for prediction and generation models.

- **Pytorch:** Pytorch is a Deep Learning framework, which is used to build models effectively by utilizing this framework. This is an alternative framework for Tensorflow.

- **Huggingface:** Hugging face libraries are very effective in building the models particularly the Transformer models. This library has a vast amount of pre-trained Transformer that includes Large Language Models, can simply fine-tuned to our dataset and get results out of it.

Note:

We have used to some of the blogs, articles and tutorials throughout our research which were mention below,

1. Introduction to Q-Learning, Hugging face, https://huggingface.co/learn/deep-rl-course/en/unit2/introduction

2. A Hands-On Introduction to Deep Q-Learning using OpenAI Gym in Python, Analytics Vidhya, https://www.analyticsvidhya.com/blog/2019/04/introduction-deep-q-learning-python/

3. The Illustrated Transformer, blog, @JayAlammar, https://jalammar.github.io/illustrated-transformer/

4. Text Summarization using BERT, GPT2, XLNet, Medium, https://medium.com/analytics-vidhya/text-summarization-using-bert-gpt2-xlnet-5ee80608e961

5. Text summarization using NLP, Medium, https://medium.com/analytics-vidhya/text-summarization-using-nlp-3e85ad0c6349

6. Text Summarization: How To Calculate Rouge Score, Medium, https://medium.com/@eren9677/text-summarization-387836c9e178

7. Metric: rouge, Hugging face, https://huggingface.co/spaces/evaluate-metric/rouge

8. T5: a detailed explanation, Medium, https://medium.com/analytics-vidhya/t5-a-detailed-explanation-a0ac9bc53e51

References

[1]. Wang, Q.; Liu, P.; Zhu, Z.; Yin, H.; Zhang, Q.; Zhang, L. A Text Abstraction Summary Model Based on BERT Word Embedding and Reinforcement Learning. *Appl. Sci.* **2019**, *9*, 4701. https://doi.org/10.3390/app9214701.

[2]. W. Liu, Y. Gao, J. Li and Y. Yang, "A Combined Extractive With Abstractive Model for Summarization," in IEEE Access, vol. 9, pp. 43970-43980, 2021, doi: 10.1109/ACCESS.2021.3066484.

[3]. L. Avramelou, N. Passalis, G. Tsoumakas and A. Tefas, "Domain-Specific Large Language Model Finetuning using a Model Assistant for Financial Text Summarization," *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*, Mexico City, Mexico, 2023, pp. 381-386, doi: 10.1109/SSCI52147.2023.10371906.

[4]. Y. K. Atri, V. Goyal and T. Chakraborty, "Multi-Document Summarization Using Selective Attention Span and Reinforcement Learning," in *IEEE/ACM Transactions on*

Audio, Speech, and Language Processing, vol. 31, pp. 3457-3467, 2023, doi: 10.1109/TASLP.2023.3316459.

[5]. D. Su, D. Su, J. M. Mulvey and H. V. Poor, "Optimizing Multidocument Summarization by Blending Reinforcement Learning Policies," in *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 3, pp. 416-427, June 2023, doi: 10.1109/TAI.2022.3201807.

[6]. H. Fei, Y. Zhang, Y. Ren and D. Ji, "Optimizing Attention for Sequence Modeling via Reinforcement Learning," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 8, pp. 3612-3621, Aug. 2022, doi: 10.1109/TNNLS.2021.3053633.

[7]. H. Jang and W. Kim, "Reinforced Abstractive Text Summarization With Semantic Added Reward," in *IEEE Access*, vol. 9, pp. 103804-103810, 2021, doi: 10.1109/ACCESS.2021.3097087.

[8]. Kohita, Ryosuke, et al. "Q-learning with language model for edit-based unsupervised summarization." *arXiv preprint arXiv:2010.04379* (2020).

[9]. Mao, Yuning, et al. "Multi-document summarization with maximal marginal relevance-guided reinforcement learning." *arXiv preprint arXiv:2010.00117* (2020).

[10]. Y. Keneshloo, T. Shi, N. Ramakrishnan and C. K. Reddy, "Deep Reinforcement Learning for Sequence-to-Sequence Models," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 7, pp. 2469-2489, July 2020, doi: 10.1109/TNNLS.2019.2929141.

[11]. Z. Li, Z. Peng, S. Tang, C. Zhang and H. Ma, "Text Summarization Method Based on Double Attention Pointer Network," in *IEEE Access*, vol. 8, pp. 11279-11288, 2020, doi: 10.1109/ACCESS.2020.2965575.

[12]. Kaichun Yao, Libo Zhang, Tiejian Luo, Yanjun Wu, Deep reinforcement learning for extractive document summarization, Neurocomputing, Volume 284, 2018, Pages 52-62, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2018.01.020.

[13]. A. Singh, R. Bhatia, and S. Saha (2022), "Complaint and Severity Identification from Online Financial Content", IEEE Transactions on Computational Social Systems. (FINancial Complaint CORpus (FINCROP) dataset).

[14]. Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[15]. Raffel, Colin, et al. "Exploring the limits of transfer learning with a unified text-to-text transformer." Journal of machine learning research 21.140 (2020): 1-67.

[16]. Grover, K., Kaur, K., Tiwari, K., Rupali, Kumar, P. (2021). Deep Learning Based Question Generation Using T5 Transformer. In: Garg, D., Wong, K., Sarangapani, J., Gupta, S.K. (eds) Advanced Computing. IACC 2020. Communications in Computer and

Information Science, vol 1367. Springer, Singapore. https://doi.org/10.1007/978-981-16-0401-0_18.

[17]. Jordan, Scott, et al. "Evaluating the performance of reinforcement learning algorithms." International Conference on Machine Learning. PMLR, 2020.