SSN MLRG at MEDIQA-SUM 2023: Automatic Text Summarization using Support Vector Machine and RoBERTa

*Sheerin Sitara Noor Mohamed[1, *] and Kavitha Srinivasan[2]*

[1, 2] Department of CSE, Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam – 603110, India

Abstract

The tremendous increase in the quantity of text requires an improved mechanism to extract the information quickly and effectively. It is very difficult for human beings to manually extract the summary of every large documents. Moreover, the searching for relevant documents from the number of documents available, and finding relevant information is also difficult. The automatic text summarization can solve this issue by identifying the most important meaningful information in a document or set of related documents and compressing them into a shorter version preserving its overall meanings. Considering this, ImageCLEF organizes MEDIQA-SUM task which is related to automatic text summarization. This paper concentrates on Dialogue2Topic summarization, a part of MEDIQA-SUM task. For this, two models are developed using SVM with TF-IDF and RoBERTa. From the results, it has been inferred that RoBERTa models achieved an highest accuracy of 0.765.

Keywords

ImageCLEF, Summarization, Support Vector Machine, RoBERTa, Singular Value Decomposition

1. Introduction

The text summarization reduces the time and effort required to read and analyse complex and lengthy documents. Moreover, the summarization gives the clear picture about the overall inference about the document in short duration of time. The text summarization plays a major role in medical field because it saves the quality time and provides future references for doctor and clinicians. The applications of text summarization include, Diaogue2Note summarization, Dialogue2Topic summarization, patient-specific text summarization, data-to-text summarization of patient records, etc. Hence, ImageCLEF [1] contribute to the text summarization field by conducing the MEDIQA-SUM task in this year. The MEDIQA-SUM task concentrates on Diaogue2Note summarization and Dialogue2Topic summarization task [2].

The Dialogue2Topic summarization task can be performed using machine learning techniques, deep learning techniques, transformer based pre-trained models. The different machine learning

[1]CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece
EMAIL: sheerinsitaran@ssn.edu.in (A. 1); kavithas@ssn.edu.in (A. 2)
ORCID: 0000-0003-1752-2107 (A. 1); 0000-0003-3439-2383 (A. 2)

techniques are, Support Vector Machine [3], Random forest [4] and decision tree [5]. The different deep learning techniques are Multi-Layer perceptron [6] and ConceptNet [7]. The different transformer based pre-trained models are BERT[8] and RoBERTa [9]. Among these techniques, SVM and RoBERTa are used in this paper for text classification because of the following reasons. They are, (i). Able to handle high dimensional data and ability to perform better with small datasets (ii). Generte representative word embedding that capture the unique characteristics of the text. Hence, in this paper, SVM and RoBERTa is used to generate classification model.

The remaining part of the paper spans across following subsections. In Sect. 2, ImageCLEF MEDIQA-SUM 2023 task and its dataset are discussed. In Sect. 3, the design of the proposed model and its implementation are explained. A brief summary about the results obtained and the performance evaluation are given in Sect. 4 with a conclusion at the end.

2. Task and Dataset Description

In this section, three sub tasks of MEDIQA-SUM 2023 and given datasets are discussed. The three sub tasks includes, Dialogue2Topic classification, Dialogue2Note Summarization and Full-Encounter Dialogue2Note Summarization.

2.1. ImageCLEF MEDIQA-SUM 2023 task

ImageCLEF, a part of Conference and Labs of the Evaluation Forum is conducting tasks related to the medical domain since 2018. The ImageCLEF MEDIQA-SUM task focuses on the automatic summarization and classification of doctor-patient conversations through three sub tasks.

In sub task A (Dialogue2Topic Classification), the topic associated with section header is identified based on the conversation snippet between a doctor and patient. The section headers will be one of twenty normalized common section labels namely, family history, history of present illness, past medical history, chief complaint, past surgical history, allergy, review of systems, medications, assessment, exam, diagnosis, disposition, plan, emergency department course, immunizations, imaging, gynecologic history, procedures, other_history and labs.

In sub task B (Dialogue2Note Summarization), the conversation between a doctor and patient is summarized and produced a clinical note section. The sub task C (Full-Encounter Dialogue2Note Summarization) is similar to sub task B. But in sub task B, given a full encounter conversation between a doctor and patient, participants are asked to generate full clinical note summarizing the conversation.

2.2. ImageCLEF MEDIQA-SUM 2023 datasets

The MEDIQA-SUM 2023 task consists of three sub tasks namely, sub task A, sub task B and sub task C. The conversational snippet , section header and section text in each sub tasks are given in Table 1. Each sub tasks consists of training set, validation set and test set.

Table 1
Dataset description for MEDIQA-SUM 2023 task

Task	Input/Output	Training Set	Validation Set	Test Set
Sub task A	**#Conversation snippet**	1201	100	198
	#Section header/ labels	20	20	-

Sub task B	**#Conversation snippet**	1201	100	200
	#Section text	1201	100	-
Sub task C	**#Conversation snippet**	67	20	40
	#Section text	67	20	-

3. System Design

The system design of the proposed models for sub task A are shown in Figure 1 and 2. This model is developed based on the conversation snippet (dialogue conversion between doctor and patient) in the training set and, the generated model is validated by predicting the label (ie., section label) for the test set.

In Figure 1, the conversation snippet and section label in the training set is tokenized into collection of words. These words are converted into a matrix of TF-IDF features using TF-IDF vectorization. This feature matrix are regularized by factoring a feature matrix into three matrices $U, \Sigma and\, V$. Then Support Vector Machine (SVM) is used to generate the classification model based on these feature vector in the training phase. Finally, the generated model is validated by predicting the section label for the conversion snippet in the test set.

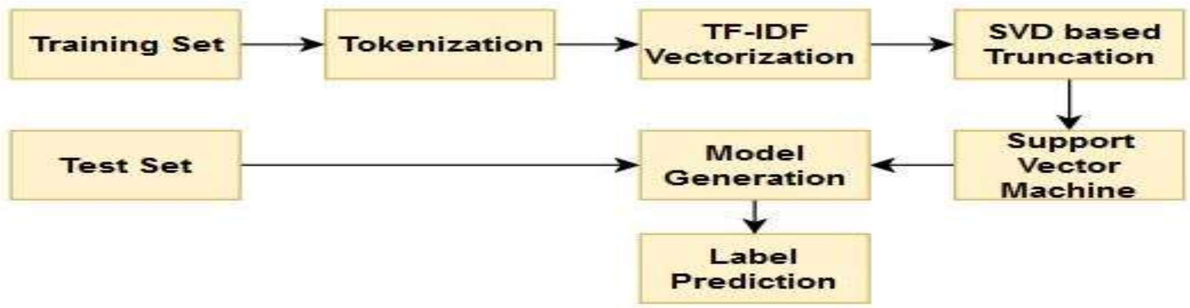

Figure 1: System design of proposed SVM model (Run 1)

In Figure 2, RoBERTa, classification model is used to predict the section label based on the conversion snippet. As RoBERTa is an pre-trained transformer based model, it supports tokenization, truncation and attention mask generation. The attention mask generation maintains the position of the padded indices so model understands which tokens should be attended. Based on the RoBERTa classification model, the model is generated for training set and validated for test set. The proposed architecture of RoBERTa is given in Figure 3. This architecture shows that RoBERTa processes the word, mask and type with respect to the ID. In RoBERTa architecture, last layer is freezes and appended with dropout, flatten and dense layer.

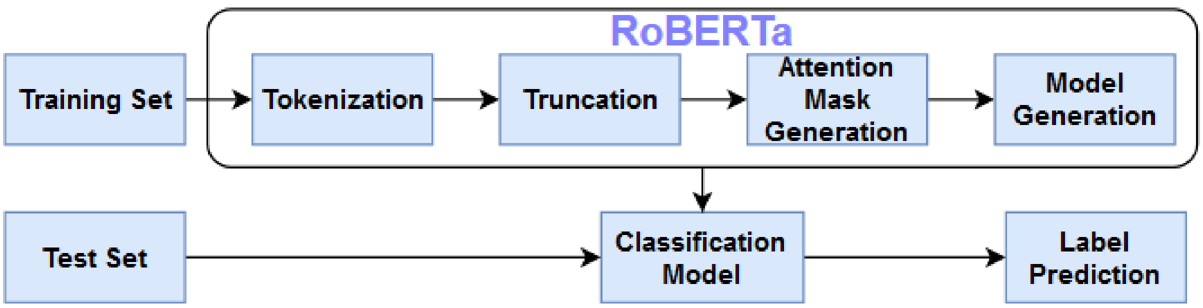

Figure 2: System design of proposed RoBERTa model (Run 2)

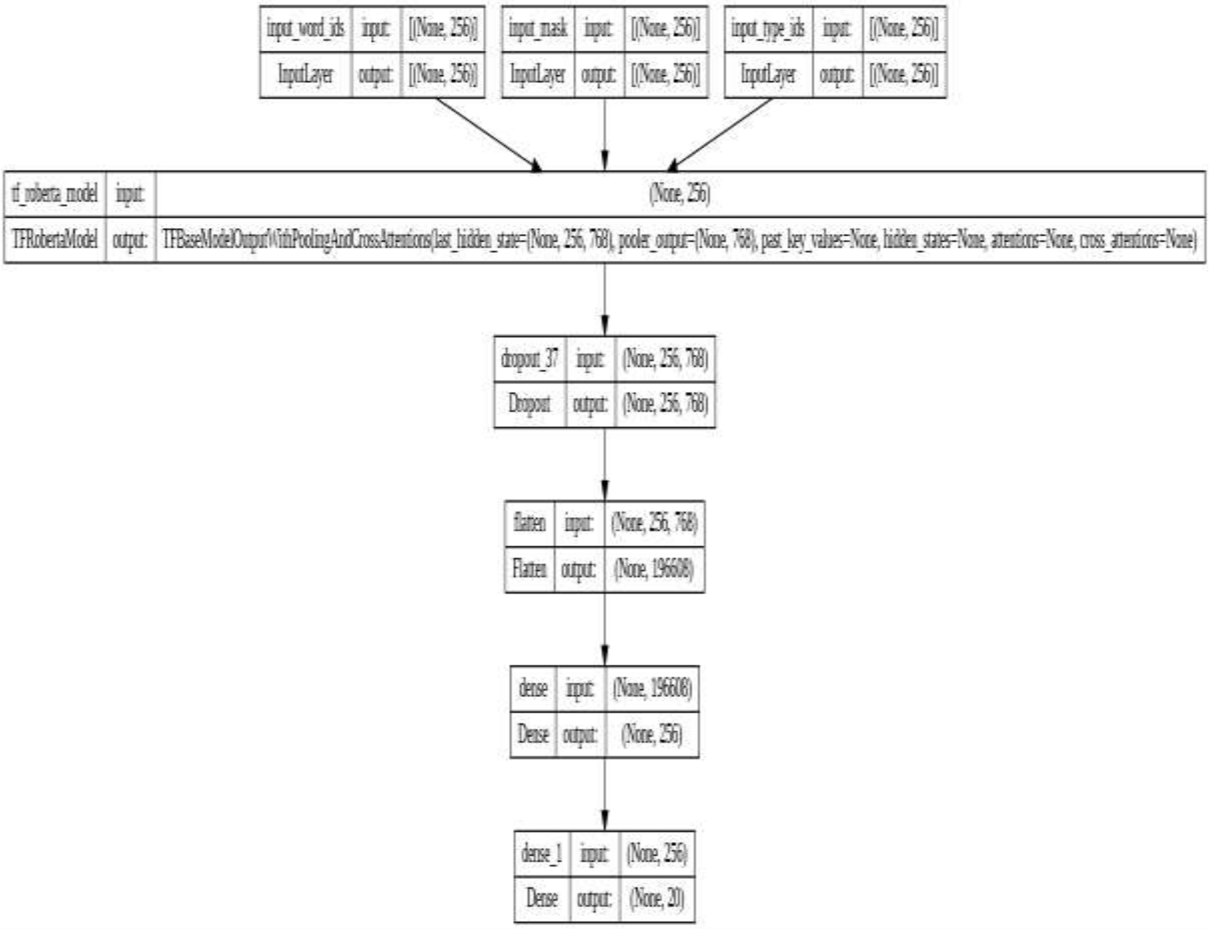

Figure 3: Proposed RoBERTa system architecture

4. Experiments and Results

The hardware and software required for the implementation of concept detection and caption prediction model includes, (i). Intel i5 processor with NVIDIA graphics card, 4800M at 4.3GHZ clock speed, 16GB RAM, Graphical Processing Unit and 2TB disk space, (ii). Linux – Ubuntu 20.04 operating system, Python 3.7 package with required libraries like tensorflow, torch, sklearn, nltk, pickle, pandas, etc.

The proposed model is executed on Dialogue2Topic Classification dataset and the performance is analyzed using three different runs as given in Table 2, are: (i). SVM along with TF-IDF and truncated SVD for the given dataset (ii). RoBERTa classification model for three epochs (iii). Same as (ii) for eight epochs.

Table 2
Brief description about each runs

Run Number	Accuracy	Run rank	Code Status
1	0.140	23	1
2	***0.740***	***7***	***1***
3	0.735	8	1

Among the results, run2 obtained better performance value in terms of accuracy and it is italicized in Table 2. From run2, the early stopping with a patience of three epochs also achieved better performance value than eight epochs. The code status of all these runs are one, which means code runs successfully and reproduces the similar results.

In Dialogue2Topic classification task, totally 11 teams were participated and 23 submissions were recorded. Among these, we have submitted seven successful submissions and achieved fourth rank among the participated teams and seventh rank in ImageCLEF 2023 MEDIQA-SUM task. The overall ranking achieved by top teams are listed in Table 3.

Table 3
Top ranking of MEDIQA-SUM 2023 task

Team Name	Accuracy	Run Rank	Team rank	Code Status
Cadence	0.820	1	1	4
HuskyScribe	0.815	2	2	5
Trendence	0.800	3	3	1
SSNSheerinKavitha	***0.765***	***7***	***4***	***1***
SuryaKiran	0.735	8	5	1
SSNdhanyadivyakavitha	0.720	10	6	1
Ds4dh	0.710	11	7	1
Uetcorn	0.710	11	8	1
SKKU-DSAIL	0.700	13	9	1
StellEllaStars	0.675	14	10	1
MLRG-JBTTM	0.665	21	11	2

5. Conclusion

This paper describes an approach to solve the ImageCLEF MEDIQA-SUM 2023 task. In this task, we have developed two models for Dialogue2Topic summarization tasks using SVM with TF-IDF (run1) and RoBERTa (run2 and run 3). From the runs, it has been inferred that RoBERTa gives better results of 0.765 which lacks only by 5.5% interms of accuracy. In the future work, the performance can be improved by incorporating medical related transformer based models. The overall performance can be improved by reducing the number of irrelevant dialogue in the conversation snippet, maintaining the number of epochs and minimum learning rate.

6. Acknowledgements

Our profound gratitude to Sri Sivasubramaniya Nadar College of Engineering, Department of CSE, for allowing us to utilize the High Performance Computing Laboratory and GPU Server for the execution of this challenge successfully.

7. References

[1] Bogdan Ionescu, Henning Müller, Ana-Maria Drăgulinescu, Wen-wai Yim, Asma Ben Abacha, Neal Snider, Griffin Adams, Meliha Yetisgen, Johannes Rückert, Alba García Seco de Herrera, Christoph M. Friedrich, Louise Bloch, Raphael Brüngel, Ahmad Idrissi-Yaghir, Henning Schäfer, Steven A. Hicks, Michael A. Riegler, Vajira Thambawita, Andrea Storås, Pål Halvorsen, Nikolaos Papachrysos, Johanna Schöler, Debesh Jha, Alexandra-Georgiana Andrei, Ahmedkhan Radzhabov, Ioan Coman, Vassili Kovalev, Alexandru Stan, George Ioannidis, Hugo Manguinhas, Liviu-Daniel Ștefan, Mihai Gabriel Constantin, Mihai Dogariu, Jérôme Deshayes, Adrian Popescu, Overview of the ImageCLEF 2023: Multimedia Retrieval in Medical, Social Media and Recommender Systems Applications, in Experimental IR Meets Multilinguality, Multimodality, and Interaction.Proceedings of the 14th

International Conference of the CLEF Association (CLEF 2023), Springer Lecture Notes in Computer Science LNCS, Thessaloniki, Greece, September 18-21, 2023.

[2] Wen-wai Yim and Asma Ben Abacha, Neal Snider, Griffin Adams and Meliha Yetisgen, Overview of the MEDIQA-Sum Task at ImageCLEF 2023: Summarization and Classification of Doctor-Patient Conversations, in: CLEF 2023 Working Notes, CEUR Workshop Proceedings, Thessaloniki, Greece, September 18-21, 2023.

[3] Begum, N., Fattah, M. A., & Ren, F. (2009). Automatic text summarization using support vector machine. international journal of innovative computing information and control, 5(7), 1987-1996.

[4] Villata, S. (2020, December). Using argument mining for legal text summarization. In IOS Press (Vol. 334, p. 184).

[5] Gambhir, M., & Gupta, V. (2022). Deep learning-based extractive text summarization with word-level attention mechanism. Multimedia Tools and Applications, 81(15), 20829-20852.

[6] Xu, Y., Vuckovic, D., Ritchie, S. C., Akbari, P., Jiang, T., Grealey, J., and Inouye, M. (2022). Machine learning optimized polygenic scores for blood cell traits identify sex-specific trajectories and genetic correlations with disease. Cell Genomics, 2(1), 100086.

[7] Patil, S., Chavan, L., Mukane, J., Vora, D., & Chitre, V. (2022). State-of-the-Art Approach to e-Learning with Cutting Edge NLP Transformers: Implementing Text Summarization, Question and Distractor Generation, Question Answering. International Journal of Advanced Computer Science and Applications, 13(1).

[8] Bitar, H., Babour, A., Nafa, F., Alzamzami, O., & Alismail, S. (2022). Increasing women's knowledge about HPV using BERT text summarization: An online randomized study. International Journal of Environmental Research and Public Health, 19(13), 8100.

[9] Cheng, R., Smith-Renner, A., Zhang, K., Tetreault, J. R., & Jaimes, A. (2022). Mapping the design space of human-ai interaction in text summarization. arXiv preprint arXiv:2206.14863.