Real-time Aspect Sentiment Quad Predictions for Vietnamese Gameshow Comments on Youtube

Trung Phan Le Chi
Faculty of Information Science and Engineering
The University of Information Technology
21522725 @uit.edu.vn

Man Nguyen Tran
Faculty of Information Science and Engineering
The University of Information Technology
21522325@uit.edu.vn

Hao Nguyen Phuc
Faculty of Information Science and Engineering
The University of Information Technology
21522047@uit.edu.vn

Duy Do Phu
Faculty of Information Science and Engineering
The University of Information Technology
21520205@uit.edu.vn

***Abstract*—Reality shows have become an indispensable part of many people's lives, they brings joy to many people. Some researchers have tried to use machine learning to exploit this data to apply to real life. In many studies, before NLP models have been used to experiment but traditional sentiment analysis can not detect the aspect of the entity, they can not optimize and exploit the full value contained in the data and Aspect-based sentiment analysis appeared. Aspect-based sentiment analysis has become popular in recent years and as an instinct humans are always searching for a challenge to improve what they have. In recent years, researchers have created a new approach to Aspect-based sentiment analysis called Aspect sentiment quad prediction. Existing studies usually consider the detection of partial sentiment elements, instead of predicting the four elements in one shot including the aspect category, aspect term, opinion term, and sentiment polarity. In this paper, we present a data set for the Aspect sentiment quad prediction task about one of the most popular shows in Vietnam called Rap Viet.**

***Index Terms*—Deep Learning, NLP, ASBA, ASQP, Apache Spark, Apache Kafka, Spark Streaming**

I. Introduction

The increase in the quantity of reality shows that make a lot of data about social media. This data is precious for many sides. The producers want to use this to improve their show. The viewer wants to use this data to help them decide whether the show is worth watching or not. Besides, this data also includes controversies about many aspects and many entities, it is a promising land for ABSA.

Sentiment analysis (SA) is a process that classifies whether data is Positive, Negative, or Neutral to help the user gain more insight into the data. Traditional sentiment analysis can not analyze the details of the specific aspects, components, parts, or functionalities of the entity. To do that, some researchers have presented a new model called aspect-based sentiment analysis that can analyze the entity's details.

As a fine-grained opinion mining problem, aspect-based sentiment analysis (ABSA) aims to analyze sentiment information at the aspect level. Typically, in ABSA there are 4 elements involved: **aspect category** detect and category the aspect that is concerned to the entity, **aspect term** which can be either explicitly or implicitly mentioned in the given text, **opinion term** describe the opinion in the text that concerned to the aspect of the entity and **sentiment polarity** denoting the sentiment class.

Lately, Aspect-based Sentiment Analysis has attracted a lot of attention with its technical approaches. Despite the popularity of ABSA, most ABSA models only attempt to perform partial prediction instead of providing a complete aspect-level sentiment picture. To do that researchers have found a new approach to ABSA called Aspect sentiment quad prediction (ASQP) aiming to predict all (aspect category, aspect term, opinion term, sentiment polarity) quads for a given opinionated sentence

In this paper:

Input: a comment in Rap Viet video on YouTube.

Output: a quad including the aspect category, aspect term, opinion term, and sentiment polarity

Nowadays, Vietnam has many reality shows and these shows usually have a lot of controversies about many entities like people, show content, music, and images ... .And most of the shows do not have a rating system to rate the aspect. We believe that if producers apply Aspect sentiment quad prediction will help them improve the quality of the show and also give viewers an objective perspective about the show, the candidates, the music, the examiner...

II. Relative works

ABSA was introduced as a SemEval task in 2014 (SE-ABSA14) providing benchmark datasets of English reviews and a common evaluation framework (Pontiki et al., 2014 [1]); the datasets were annotated with aspect terms (e.g. "hard disk", "pizza") and their polarity for laptop and restaurant reviews, as well as coarser aspect categories (e.g., FOOD) and their polarity only for the restaurant's domain. The task was repeated in SemEval 2015 (SE-ABSA15) aiming to facilitate more in-depth research by providing a new ABSA

Identify applicable funding agency here. If none, delete this.

Aspect category	Definition
Candidate voice	Comments about the candidates' voice, the way they sing
Candidate flow	Comments mention the candidates' flow, a technic in rap
Candidate dancing	Comments about how they dance on the stage
Candidate general	Comments describe other aspects of the candidate (outlook, style, character...)
Examiner general	Comments about the examiners
Show stage	Comments about how the stage look
Show general	About other aspect of the show
Music	Comments refer to the music, the melody, the song
Others	Spam comments

TABLE I: Aspect category definition

framework in which all the identified constituents of the expressed opinions (aspects, opinion target expressions, and sentiment polarities) meet a set of guidelines/specifications and are linked to each other within tuples. In the context of the new framework, an aspect category is defined as a combination of an entity type E (e.g. LAPTOP, KEYBOARD, CUSTOMER SUPPORT, RESTAURANT, FOOD) and an attribute type A (e.g. USABILITY, QUALITY, PRICE) of E, making more explicit the difference between entities and the particular facets that are being evaluated (Pontiki et al., 2015 [2]). The SemEval-2016 task-5 (SE-ABSA16) (Pontiki et al., 2016 [3]) dataset extended SE-ABSA15 to new domains such as Hotels, Consumer Electronics, Telecom, Museums, and other languages.

In several recent years, many researchers have published their works about ABSA with positive results ((Nazir et al., 2020) [4], (Do et al., 2019 [5]). But most of them do not predict all quads in one shot. In 2020, (Peng et al. (2020) [6]) propose the aspect sentiment triplet extraction (ASTE) task, which has received lots of attention **(**Xu et al., 2020 [7] **;** Huang et al., 2021 [8] **;** Mao et al., 2021 [9]...**)**. In 2021, (Zhang et al. 2021 [10]) presented Aspect Sentiment Quad Prediction as a new step in NLP.

In Vietnam many universities and researchers have published their datasets, works in ABSA: Vietnamese ABSA corpus about smartphone reviews (Mai and Le, 2018 [11]), SA-VLSP2018 dataset (published by H. T. Nguyen et al. [12]) about hotel and restaurant reviews. A dataset on the same domain as VLSP was created by (Nguyen et al., 2019) for the two tasks of sentiment classification and aspect extraction that were the focus of the earlier work. In addition, (Thin et al., 2021 [13]) constructed a sentence-level dataset for the same topic that was annotated with high inter-annotator agreements in two earlier kinds of research. UIT-ViSFD benchmark dataset is created for evaluating ABSA for mobile e-commercial by (Luc Phan et al., 2021 [14]). (Thanh et al., 2021) presented the UIT-ViSD4SA dataset with span detection for ABSA, which is a benchmark Vietnamese smartphone feedback dataset. Most recently, (Tran et al., 2022 [15]) performed ABSA on a dataset about e-commercial beauty product reviews. A lot of research has been organized to study ABSA in Vietnamese but there are shortage in research about the new approaches of ABSA (Thin and Nguyen, 2023 [16]) and even more shortage about Aspect sentiment quad prediction (ASQP).

III. Data

A. Source and Crawling Method

Our dataset was collected from YouTube videos about Rap Viet season 3 by using YouTube API provided by Google and Selenium library in Python. We chose Rap Viet because it is the most popular competitive reality show in Vietnam, it has many controversies from the viewers about many entities that appear in the show. Our dataset includes 3033 comments crawled from a playlist including 118 videos about Rap Viet in season 3. About sentiment polarity, we design it with 3 labels: Positive, Negative, and Neutral. For the aspect category detection, we design this subtask with 8 aspects including Candidate voice, Candidate flow, Candidate general, Candidate dancing, Examiner general, Show stage, Show general and Music. We also have an aspect named Others to detect spam or comments that do not have any aspect or sentiment. Their definitions are shown in Table [I].

We split the dataset into 4 species: train set with **2110** comments, validation set with **301** comments, test set with **301** comments, and stream set with **301** comments for simulating data streaming with Spark and Kafka.

B. Annotation Process

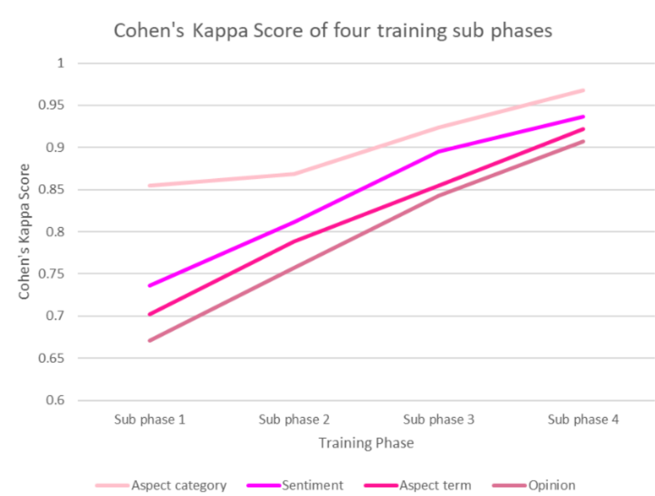

Fig. 1: Conhen's Kappa Score

First of all, we define a basic annotation guideline. For training, we randomly take 100 comments to annotate, then we

Set	Comments	Avg aspect/comment	Positive	Negative	Neutral	Total
Train	2110	1.201895735	1494	130	912	2536
Dev	301	1.219269103	230	20	117	367
Test	301	1.215946844	213	23	130	366
Stream	301	1.146179402	203	13	129	345

TABLE II: The overview statistics of the dataset

calculate Cohen's Kappa score for those annotated comments. For labels that do not have enough 4 agreements, we gather and discuss a new rule to update our guidelines. We trained four rounds to obtain Cohen's Kappa score higher than 90% before performing data annotation independently. The result of each round is shown in Figure [1]. After the training phase, the rest of the data was split into 4 parts and each member annotated 1 part following the rules in our annotation guideline.

C. Statistics

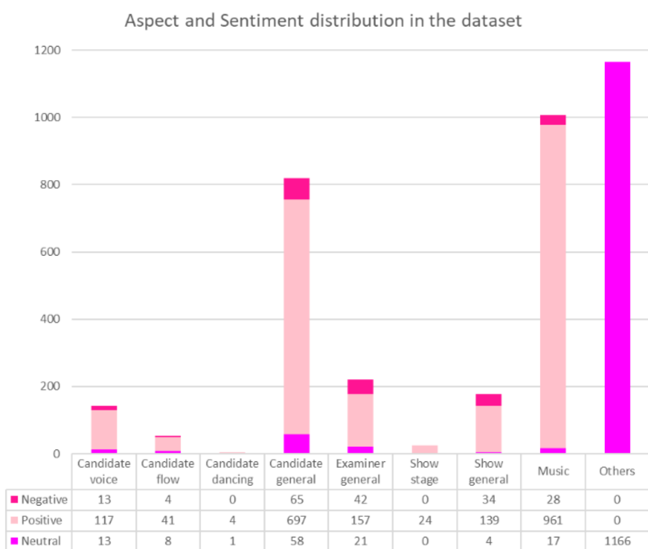

Fig. 2: Aspect category and Sentiment polarity distribution in the dataset

Our dataset contains 3013 comments, including 8 sentiment aspects except "Others", with each sentiment aspect having three sentiment labels: Neural, Positive, and Negative. The overview statistics of the training, validation, and testing set are shown in Table [II]. As you can see in Figure [2], there are 3 aspects that receive a high number of comments compared to other aspects: "Others" with 1166 comments classified, "Music" (1006) and "Candidate general" (820). On the other hand, we have identified two aspects, "Show stage" and "Candidate dancing", each exhibiting a lower count compared to other aspects present. Aspect "Show stage" comprises 24 comments, while Aspect "Candidate dancing" has 5. The presence of these low counts poses inherent challenges for our sentiment analysis models. In terms of sentiment distribution, Positive labels dominate across all aspects (accounting for from 71% to 100%), showing more interest of viewers to rappers and shows. This also shows that there is an imbalance in the number of labels in each aspect, which can affect the performance of deep learning models. Overall, this dataset provides a valuable foundation for sentiment analysis, with further exploration recommended to uncover deeper insights into specific aspects and sentiments.

D. Data Preprocessing

For data preprocessing, we have applied many techniques:
- Remove HTML
- Convert unicode, normalize acronyms, teencode, emoji
- Remove unnecessary, duplication characters

Padding is used with a maximum length of 128, which is chosen based on the sentence length distribution chart of the dataset shown in Figure [3].

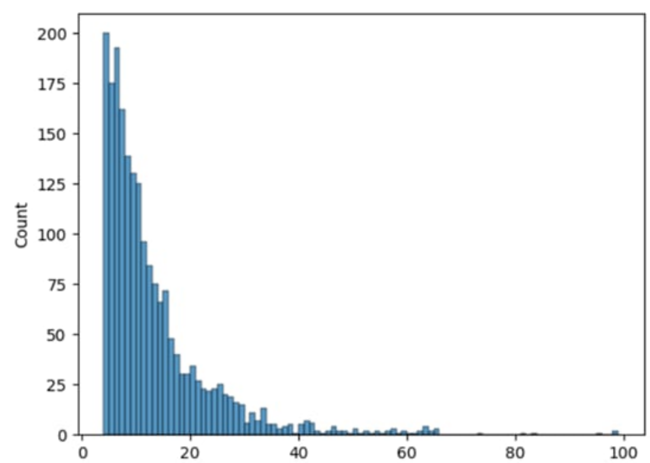

Fig. 3: Sentence length distribution of the dataset

IV. Model

A. Approach

We use the Paraphrase Generation method presented in Aspect Sentiment Quad Predictions as a Paraphrase Generation paper (Zhang et al., 2021 [10]). With an input comment, aspect sentiment quad predictions (ASQP) aim to predict aspect category, aspect term, opinion term, and sentiment polarity. This method paraphrases the input comment to neglect unnecessary details and highlight the major sentiment elements:

Aspect category là **Sentiment polarity** bởi vì **Aspect term** là **Opinion term**

Two examples of paraphrasing the input comments for training using the Paraphrase Generation Method are presented in Figure [4]. After training the model with these paraphrased

Input – 1	*Double2T đỉnh quá anh em*
Label – 1 ⇓ Target – 1	(c, a, o, p) (Candidate general, Double2T, đỉnh, POS) ⇓ Candidate general là tuyệt bởi vì Double2T là đỉnh
Input – 2	*Rap việt mùa này mất chất rồi*
Label – 1 ⇓ Target – 1	(c, a, o, p) (Show stage, Rap việt, mất chất, NEG) ⇓ Show stage là tệ bởi vì Rap việt là mất chất

Fig. 4: Two examples for Paraphrase Generation Method

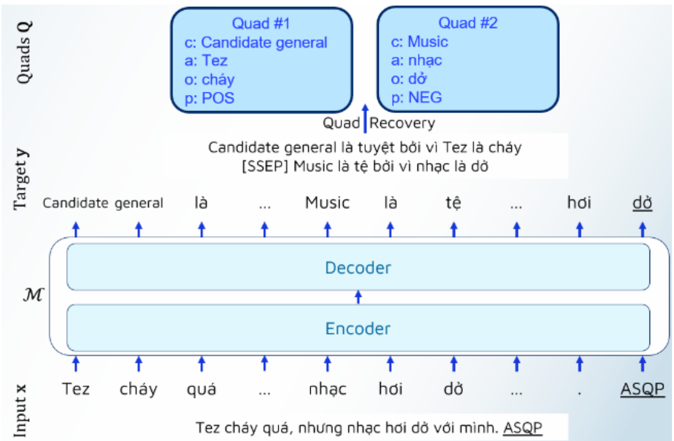

Fig. 5: Paraphrase Generation model framework overview

sentences, it will perform the prediction as the form above, which can be used to extract the sentiment quads. The details of this process are shown in Figure [5].

B. Model

For Vietnamese text generation, we use two models: ViT5 and BARTPho

Text generation is a process where a model produces written content, imitating human language patterns and styles. The process involves generating coherent and meaningful text that resembles natural human communication. Text generation has gained significant importance in various fields, including natural language processing, content creation, customer service, and coding assistance.

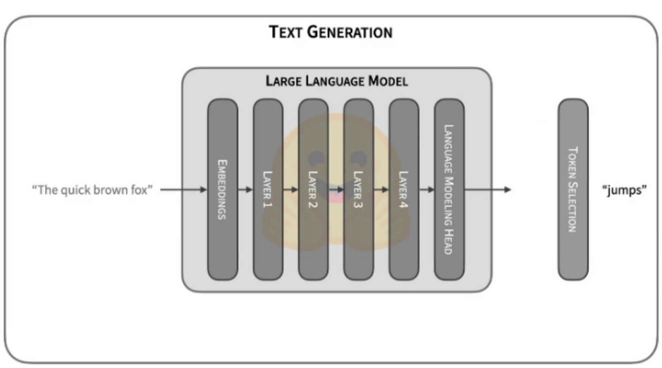

Fig. 6: Example of Text Generation by Large Language Model

1) ViT5: T5 is an encoder-decoder model pre-trained on a multi-task mixture of unsupervised and supervised tasks for which each task is converted into a text-to-text format. T5 works well on a variety of tasks out-of-the-box by prepending a different prefix to the input corresponding to each task

ViT5 is a pretrained sequence-to-sequence Transformer model for the Vietnamese language. It is a pretrained Transformer-based encoder-decoder model for the Vietnamese language. With T5-style self-supervised pretraining, ViT5 is trained on a large corpus of high-quality and diverse Vietnamese texts.

2) BARTPho: BART is a denoising auto encoder for pretraining sequence-to-sequence models. It uses a standard Transformer-based neural machine translation architecture which, despite its simplicity, can be seen as generalizing BERT (due to the bidirectional encoder), GPT (with the left-to-right decoder), and many other more recent pretraining schemes.

BARTPho is a large-scale monolingual sequence-to-sequence model pre-trained for Vietnamese. BARTpho uses the "large" architecture and the pre-training scheme of the sequence-to-sequence denoising autoencoder BART, thus it is especially suitable for generative NLP tasks.

V. EXPERIMENT

A. Experiment settings

Because the method used in this paper uses a text generation model to paraphrase the input comments, we will apply two models in two ways: No preprocess and Preprocess the dataset to check the impact of data preprocessing on models' ability to learn and generate text.

Adam with 3e-4 learning rate and 1e-8 epsilon is used as Optimizer. We use batch size of 16 and 20 epochs with an Early Stopping function.

B. Evaluation Metrics

A sentiment quad prediction is counted as correct if and only if all the predicted elements are exactly the same as the gold labels (Zhang el al [10]). F1-score macro, which is the harmonic mean of the precision and recall, is used because of the imbalance of aspect and sentiment in the dataset. It thus symmetrically represents both precision and recall in one metric and gives us a more general view of the models' performance. The formula of F1-score is as follows with TP, FP, and FN denoted for True Positive, False Positive and False Negative respectively:

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1 = \frac{2*Precision*Recall}{Precision+Recall} = \frac{2*TP}{2*TP+FP+FN}$$

VI. RESULTS AND DISCUSSION

Our model was successful in identifying and extracting important aspects from the input text. We have divided the results into 3 tables with different evaluation aspects using the F1-score index.

		Precision	Recall	F1-macro
No preprocess	ViT5	0.2988	0.3470	0.3211
	BARTpho	0.2812	0.2650	0.2729
Preprocess	ViT5	0.2909	0.3060	0.2983
	BARTpho	0.3084	0.2923	0.3001

TABLE III: The overall experimental results

	No preprocess		Preprocess	
	ViT5	BARTpho	ViT5	BARTpho
Candidate voice	0.5263	0.3750	0.2222	0.5263
Candidate flow	0.4	0	0.2857	0
Candidate dancing	0	0	0	0
Candidate general	0.3896	0.5333	0.4379	0.4444
Examiner general	0.3	0.1667	0.3514	0.1667
Show stage	0	0.2857	0.4	0.4
Show general	0.2326	0.1579	0.4324	0.16
Music	0.5394	0.6504	0.6479	0.6139
Others	0.7348	0.7085	0.6932	0.6496

TABLE IV: Aspect Category Detection F1-macro

	No preprocess		Preprocess	
	ViT5	BARTpho	ViT5	BARTpho
Negative	0.2222	0.2069	0.3333	0.1429
Positive	0.5923	0.7196	0.6440	0.6277
Neutral	0.7174	0.68	0.6767	0.6667

TABLE V: Sentiment Classification F1-macro

In the overall experimental results shown in Table [III], we evaluate the performance of ViT5 and BARTpho pre-trained models on two methods, with and without data preprocessing. Without preprocessing, ViT5 exhibited superior precision (0.2988), recall (0.3470), and F1-macro (0.3211) compared to BARTpho, which showed lower scores across all metrics (precision: 0.2812, recall: 0.2650, F1-macro: 0.2729). However, when data preprocessing was applied, the dynamics shifted. ViT5's performance saw a slight decline in precision, recall, and F1-macro (0.2909, 0.3060, 0.2983, respectively), while BARTpho showcased notable improvements (precision: 0.3084, recall: 0.2923, F1-macro: 0.3001). Interestingly, ViT5 consistently maintained higher recall in both scenarios. These findings suggest that the impact of data preprocessing varies between models, with ViT5 showcasing robust recall and BARTpho benefiting more from preprocessing in terms of precision and F1-macro. The observed results prompt consideration of the nuances in model behavior, emphasizing the importance of tailoring preprocessing strategies to specific models and data characteristics.

Next, Table [IV] highlights notable aspects where the models encountered challenges in predicting, such as "Candidate flow" with the Bartpho model, "Show stage" with the ViT5 model in the absence of data preprocessing, and "Candidate dancing" for both models. These instances underscore the intricacies and nuances in aspect prediction that the models grapple with. In the absence of data preprocessing, a nuanced comparison revealed that the Vit5 model outperforms Bartpho in aspects like "Candidate voice," "Examiner general," "Show general," and "Others." On the other hand, Bartpho demonstrated superior performance in the remaining aspects. However, with preprocessed data, the dynamics shifted, and Vit5 exhibited better performance in aspects such as "Examiner general," "Show general," "Music," and "Others," whereas Bartpho excelled in the opposite set of aspects. A noteworthy observation is that, following data preprocessing, Vit5 successfully identified "Show stage" as an aspect with an F1-score of 0.4, achieving parity with the performance of Bartpho for this specific aspect. These findings underscore the impact of data preprocessing on aspect category detection, emphasizing the need for tailored approaches to optimize model performance across diverse aspects in these tasks.

Finally in Table [V], like the previous approaches, we evaluate the performance of the two models for sentiment classification. Without preprocessing, ViT5 demonstrated improvements in all sentiment categories, particularly in Negative and Neutral sentiments, while BARTpho showcased robust Positive sentiment predictions but a decrease in Negative and Neutral sentiments. Upon applying data preprocessing, ViT5's Positive sentiment predictions improved notably, but there was a decline in predicting Neutral sentiments. Conversely, BARTpho's performance degraded across all sentiment categories, with a significant decrease in Negative sentiment prediction. These findings underscore the nuanced impact of data preprocessing on model performance, providing valuable insights into the strengths and weaknesses of each model. The report suggests that ViT5 benefits from preprocessing, particularly in enhancing Positive sentiment predictions, while BARTpho's performance is more sensitive to data preprocessing, requiring further investigation for potential improvements.

In conclusion, although we successfully applied the Paraphrase method on a Vietnamese dataset, the results that we achieved are still pretty low due to the tiny dataset, which only contains 3013 comments and the highly imbalance of the aspect category and sentiment polarity labels.

A. Error analysis

We have found that both models have five types of errors cannot detect the aspect/opinion term, detect the wrong aspect/opinion term, misclassify the sentiment polarity, cannot detect the aspect category, and detect the wrong aspect category. As you can see in the first comment of Table [VI], which is several predictions of the ViT5 model, 'quả beat đã thật' is supposed to be ('Music', 'beat', 'Positive', 'đã') but the model cannot detect all the labels. In the second comment, the model detect wrong range of Aspect and Opinion term, from 'Nhận xét của anh Thai VG' to 'Thai VG' and from 'lạc nhịp' to 'có vẻ hơi lạc nhịp'. For the third sentence, the model misclassified the sentiment polarity from Neutral to Positive. In the final example, the Aspect category should be both Candidate general but Others was predicted by the model.

Reviews	Aspect category	Aspect term	Sentiment polarity	Opinion term
quả beat đã thật	NULL	NULL	NULL	NULL
Nhận xét anh Thai VG có vẻ hơi lạc nhịp	Examiner general	Thai VG	Negative	có vẻ hơi lạc nhịp
Nói thật chứ xem lại màn trình diễn của 24K.Right và Minh Lai quá thật out trình vòng chinh phục	Others Others	24K.Right và Minh Lai Minh Lai	Neutral Neutral	NULL NULL

TABLE VI: Some incorrect prediction examples of ViT5

VII. Spark Streaming

Spark Streaming is a real-time data processing framework within the Apache Spark ecosystem, designed for seamless integration with batch processing. It enables high-throughput, fault-tolerant stream processing, allowing users to process live data streams in micro-batches.

Spark Streaming can provide fault recovery and scalable, parallel processing. With its flexibility and compatibility with various data sources, Spark Streaming is a powerful tool for handling continuous data streams in a distributed and resilient manner.

Our system is divided into two parts: real-time data collecting and real-time analysis. Its architecture is presented in the figure below:

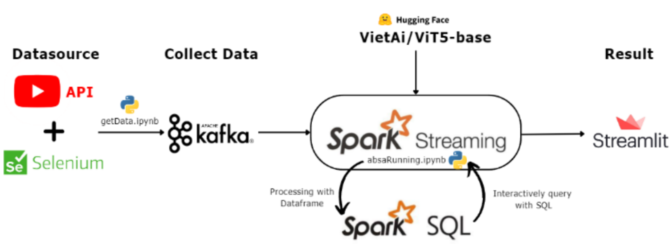

Fig. 7: The end-to-end system's architecture using Spark Streaming

A. Real-time Data Collecting

For the first part, we're tapping into the YouTube API to gather data. This valuable information is then sent off to a Kafka producer, creating a smooth flow for data streaming.

Once in Kafka, our consumer component steps in to grab the data. This data journey continues as it makes its way into Spark Streaming DataFrames. This setup allows us to process the data in small, continuous chunks, enabling us to keep an eye on sentiment changes in real time.

B. Real-time Analysis

In this part, we receive the data from the TCP socket and apply the data preprocessing layer that we mentioned in Section [III]. Result of this process is presented in Figure [8]. Then, we deploy the ViT5 pre-trained model which is trained offline. After getting the results, we categorize them into for each Vietnamese Rap contestant from aspect terms that are predicted, we create statistical charts and deploy a website application using Streamlit based on those aspect categories shown in Figure [9]. This website will help users catch up with the overall statistics of Rap Viet candidates by knowing their aspect category and sentiment polarity that are commented on livesream YouTube videos by audiences.

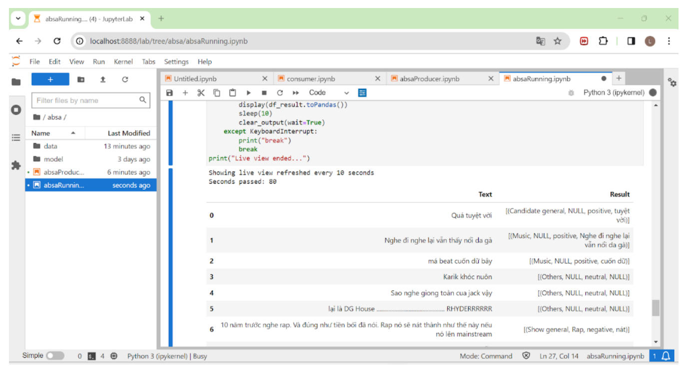

Fig. 8: Consumer results

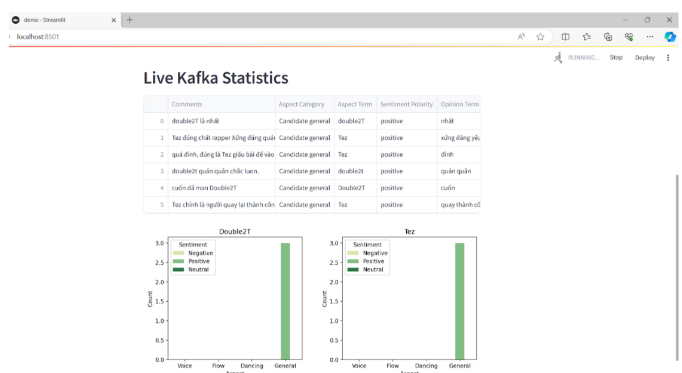

Fig. 9: Real time candidate statistics Streamlit demo

VIII. Conclusion and Future Work

In this paper, we have presented a new dataset for ABSA about one of the most popular game shows in Vietnam for quad predictions: aspect category, aspect term, opinion term, and sentiment polarity. We applied the Paraphrase Generation method to this dataset, with the aim of predicting a quad that includes all 4 elements in one shot. We deployed two models

ViT5 and BARTPho to our dataset and received an acceptable result.

In the future, we plan to collect more data and aim to make our dataset more balanced and even more we want to expand our topic to more shows or even to other fields like education, social media KOLs.... In our ideal, we want to build a system that can automatically collect, preprocess, analyze the comments, and visualize the result, which will be useful in many ways. If it is applied in shows, it will help the producer have new ideas, and help them improve the show in many aspects, and with the customer, they will know the overall candidates' statistics. Further, we plan to research more about ASQP, multimodel ABSA, and try LLAMA in order to develop a Social Listening system.

REFERENCES

[1] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, P. Nakov and T. Zesch, Eds. Dublin, Ireland: Association for Computational Linguistics, Aug. 2014, pp. 27–35. [Online]. Available: https://aclanthology.org/S14-2004

[2] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "SemEval-2015 task 12: Aspect based sentiment analysis," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, P. Nakov, T. Zesch, D. Cer, and D. Jurgens, Eds. Denver, Colorado: Association for Computational Linguistics, Jun. 2015, pp. 486–495. [Online]. Available: https://aclanthology.org/S15-2082

[3] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, and G. Eryiğit, "SemEval-2016 task 5: Aspect based sentiment analysis," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, S. Bethard, M. Carpuat, D. Cer, D. Jurgens, P. Nakov, and T. Zesch, Eds. San Diego, California: Association for Computational Linguistics, Jun. 2016, pp. 19–30. [Online]. Available: https://aclanthology.org/S16-1002

[4] A. Nazir, R. Yuan, L. Wu, and L. Sun, "Issues and challenges of aspect-based sentiment analysis: A comprehensive survey," *IEEE Transactions on Affective Computing*, vol. PP, pp. 1–1, 01 2020.

[5] H. H. Do, P. W. Prasad, A. Maag, and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: a comparative review," *Expert systems with applications*, vol. 118, pp. 272–299, 2019.

[6] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, pp. 8600–8607, Apr. 2020. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/6383

[7] L. Xu, H. Li, W. Lu, and L. Bing, "Position-aware tagging for aspect sentiment triplet extraction," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds. Online: Association for Computational Linguistics, Nov. 2020, pp. 2339–2349. [Online]. Available: https://aclanthology.org/2020.emnlp-main.183

[8] L. Huang, P. Wang, S. Li, T. Liu, X. Zhang, Z. Cheng, D. Yin, and H. Wang, "First target and opinion then polarity: Enhancing target-opinion correlation for aspect sentiment triplet extraction," 2021.

[9] Y. Mao, Y. Shen, C. Yu, and L. Cai, "A joint training dual-mrc framework for aspect based sentiment analysis," 2021.

[10] W. Zhang, Y. Deng, X. Li, Y. Yuan, L. Bing, and W. Lam, "Aspect sentiment quad prediction as paraphrase generation," 2021.

[11] L. Mai and H. B. Le, "Aspect-based sentiment analysis of vietnamese texts with deep learning," in *Asian Conference on Intelligent Information and Database Systems*, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:3750939

[12] H. T. M. Nguyen, H. V. Nguyen, Q. T. Ngo, L. X. Vu, V. M. Tran, B. X. Ngo, and C. A. Le, "Vlsp shared task: Sentiment analysis," *Journal of Computer Science and Cybernetics*, vol. 34, no. 4, p. 295–310, Jan. 2019. [Online]. Available: https://vjs.ac.vn/index.php/jcc/article/view/13160

[13] D. Thin, N. Nguyen, T. Truong, L. Le, and D. Vo, "Two new large corpora for vietnamese aspect-based sentiment analysis at sentence level," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, pp. 1–22, 05 2021.

[14] L. Luc Phan, P. Huynh Pham, K. Thi-Thanh Nguyen, S. Khai Huynh, T. Thi Nguyen, L. Thanh Nguyen, T. Van Huynh, and K. Van Nguyen, "Sa2sl: From aspect-based sentiment analysis to social listening system for business intelligence," in *Knowledge Science, Engineering and Management*, H. Qiu, C. Zhang, Z. Fei, M. Qiu, and S.-Y. Kung, Eds. Cham: Springer International Publishing, 2021, pp. 647–658.

[15] Q.-L. Tran, P. T. D. Le, and T.-H. Do, "Aspect-based sentiment analysis for Vietnamese reviews about beauty product on E-commerce websites," in *Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation*, S. Dita, A. Trillanes, and R. I. Lucas, Eds. Manila, Philippines: Association for Computational Linguistics, Oct. 2022, pp. 767–776. [Online]. Available: https://aclanthology.org/2022.paclic-1.84

[16] D. Thin and N. Nguyen, "Aspect-category based sentiment analysis with unified sequence-to-sequence transfer transformers," *VNU Journal of Science: Computer Science and Communication Engineering*, vol. 39, no. 2, 2023. [Online]. Available: //jcsce.vnu.edu.vn/index.php/jcsce/article/view/662

[17] K. N. T. Thanh, S. H. Khai, P. P. Huynh, L. P. Luc, D.-V. Nguyen, and K. N. Van, "Span detection for aspect-based sentiment analysis in Vietnamese," in *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation*, K. Hu, J.-B. Kim, C. Zong, and E. Chersoni, Eds. Shanghai, China: Association for Computational Lingustics, 11 2021, pp. 318–328. [Online]. Available: https://aclanthology.org/2021.paclic-1.34

[18] M. Hoang, O. A. Bihorac, and J. Rouces, "Aspect-based sentiment analysis using BERT," in *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, M. Hartmann and B. Plank, Eds. Turku, Finland: Linköping University Electronic Press, Sep.–Oct. 2019, pp. 187–196. [Online]. Available: https://aclanthology.org/W19-6120

[19] N. Nguyen, T. Phan, D.-V. Nguyen, and K. Nguyen, "ViSoBERT: A pre-trained language model for Vietnamese social media text processing," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, H. Bouamor, J. Pino, and K. Bali, Eds. Singapore: Association for Computational Linguistics, Dec. 2023, pp. 5191–5207. [Online]. Available: https://aclanthology.org/2023.emnlp-main.315

[20] H.-Q. Dang, D.-D.-A. Nguyen, and T.-H. Do, "Multi-task solution for aspect category sentiment analysis on vietnamese datasets," in *2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, 2022, pp. 404–409.

[21] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A survey on aspect-based sentiment analysis: Tasks, methods, and challenges," 2022.

[22] K. V. Nguyen, V. D. Nguyen, P. X. V. Nguyen, T. T. H. Truong, and N. L.-T. Nguyen, "Uit-vsfc: Vietnamese students' feedback corpus for sentiment analysis," in *2018 10th International Conference on Knowledge and Systems Engineering (KSE)*, 2018, pp. 19–24.