ABCD Team at FinancES 2023: An Unified Generative Framework for the Financial Targeted Sentiment Analysis in Spanish

Dang Van Thin[1,2,*], Dai Nguyen Ba[1,2,3], Duong Ngoc Hao[1,2] and
Ngan Luu-Thuy Nguyen[1,2]

[1] *University of Information Technology-VNUHCM, Quarter 6, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam*

[2] *Vietnam National University, Ho Chi Minh City, Vietnam*

[3] *Faculty of Information Science and Engineering, University of Information Technology-VNUHCM*

Abstract

This paper presents our participation in the IBERLEF 2023 Task - FinancES in Spanish, focusing on two sub-tasks: Financial targeted sentiment analysis and Financial Sentiment Analysis at the document level for companies and consumers. To address these sub-tasks, we propose a unified generative framework that leverages strong pre-trained language models capable of simultaneously extracting all relevant elements from financial news headlines. Additionally, we introduce two simple auxiliary tasks designed to provide the model with additional information to distinguish sentiment classes for different perspectives. The experimental results validate the effectiveness of our approach, as our participation system achieved a Top 3 ranking in both sub-tasks. Specifically, our best model achieved a result of 78.2175% and 61.0373% in terms of F1-score for Task 1 - identifying the target term and its corresponding sentiment and Task 2 - classifying the sentiments for the companies and consumers, respectively.

Keywords

Financial Targeted Sentiment Analysis, Spanish language, Unified generative framework, sentiment analysis, aspect-based sentiment analysis,

1. Introduction

The shared-task IBERLEF 2023 Task [1] - FinancES: Financial Targeted Sentiment Analysis [2] in Spanish aims to extract the targeted sentiment analysis in the field of microeconomics. In this shared task, two sub-tasks were proposed for participants. The first challenge, called the targeted sentiment task, aims to extract the target entity in the content and identify the sentiment polarity towards the target. For example, given a new headline, “Acuerdos comerciales, sinónimo de oportunidades para República Dominicana” the output of this task is “Acuerdos comerciales”

IberLEF 2023, September 2023, Jaén, Spain

[*] Corresponding author.

thindv@uit.edu.vn (D. V. Thin); 21521914@gm.uit.edu.vn (D. N. Ba); haodn@uit.edu.vn (D. N. Hao); ngannlt@uit.edu.vn (N. L. Nguyen)

https://nlp.uit.edu.vn/ (D. V. Thin); https://nlp.uit.edu.vn/ (N. L. Nguyen)

0000-0001-8340-1405 (D. V. Thin); 0009-0008-8559-3154 (D. N. Ba); 0000-0002-3136-0661 (D. N. Hao); 0000-0003-3931-849X (N. L. Nguyen)

and “positive” for the target and sentiment class, respectively. On the other hand, another task is called document-level sentiment analysis, which aims to classify the sentiment polarity for both the company and consumer aspects. Using the same input as above, the output of this task would be “positive” for both the company and consumer aspects, respectively. The sentiment class in both sub-tasks is assigned one of the “positive”, “negative”, and “neutral” values.

With the power of pre-trained generative language models such as T5 [3] and BART [4], many natural language processing tasks have been successfully addressed as text generation problems, surpassing the performance of traditional approaches, e.g. Named Entity Recognition [5], Aspect-based Sentiment Analysis [6] and etc. Therefore, this paper presents a unified generative framework that integrates two sub-tasks into a cohesive generative formulation. Subsequently, we refine and optimize the pre-trained sequence-to-sequence language models to address both sub-tasks within an end-to-end framework effectively. Moreover, we design two auxiliary tasks to provide the models regarding the relationship between the sentiment classes of the three sentiment objects.

2. Related work

The easy accessibility of financial texts has significantly increased with the widespread use of the Internet and the growing need for market transparency. As a result, the field of financial text analysis has emerged. While the concept of utilizing textual analysis in the financial markets is not entirely novel, the influence of sentiment analysis on financial markets has been firmly established.

Recently, Yıldırım et al. [7] explored different deep Learning approaches for Sentiment Analysis on the financial dataset by comparing deep learning classifiers to traditional machine learning approaches. Their findings highlighted the superior performance of LSTM models, including bidirectional LSTM and LSTM with dropout, and revealed similar success rates among various optimizers. Another relevant study by Lee et al. (2020)[8] employed the pre-trained BERT-base model and achieved high accuracy in recognizing investor sentiment after fine-tuning on a labelled sentiment dataset. Mishev et al.[9] investigated the evolution of sentiment analysis methods from lexicons to transformers, emphasizing NLP transformers’ superiority and ability to capture semantic meaning effectively. In recent years, the implementation of the RNN-LSTM network by Kohsasih et al. [10] demonstrated that this approach can show a significant improvement in sentiment analysis task. A study conducted by Ong et al. [11] explored the statistical connection between aspect-based sentiment labels and specific stocks. The research revealed a distinct correlation, particularly in relation to aspects such as inflation and the economy, highlighting their significant influence on stock prices.

Unlike previous approaches that rely on classification-based approaches for sentiment analysis on financial data, in this paper, we convert two challenges in the shared-task [2] as a generative problem and utilize the power of pre-trained language models combined with two auxiliary tasks.

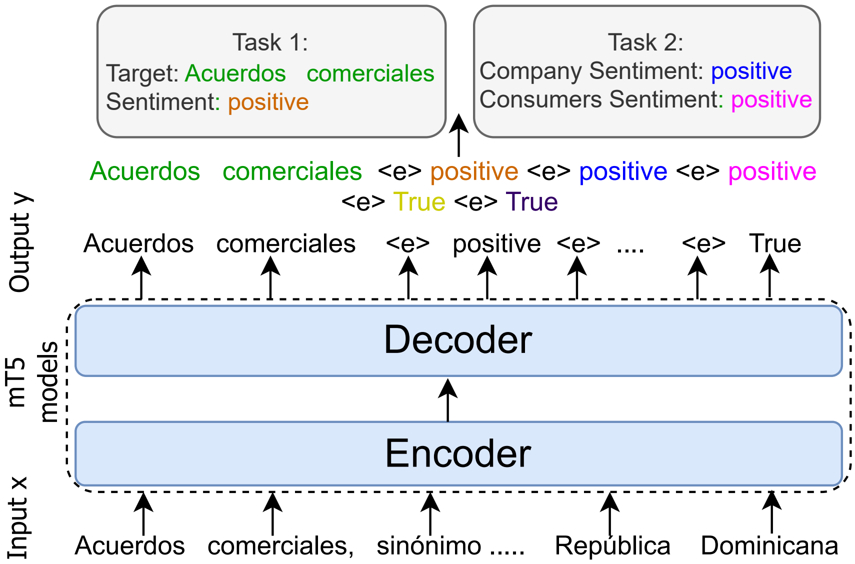
Task 1:
Target: Acuerdos comerciales
Sentiment: positive
Task 2:
Company Sentiment: positive
Consumers Sentiment: positive
Acuerdos comerciales <e> positive <e> positive <e> positive
<e> True <e> True
Output y
Acuerdos comerciales <e> positive <e> .... <e> True
mT5 models
Decoder
Encoder
Input x
Acuerdos comerciales, sinónimo ..... República Dominicana

Figure 1: Overview of our unified generative framework for the Financial targeted sentiment analysis shared-task.

3. Approach

Figure 1 shows the overview of the generative framework for both tasks. Rather than approaching the two challenges as classification problems, the tasks are formulated such that the input sentence $\boldsymbol{x}$ is processed, and the resulting output is a target sequence $\boldsymbol{y}$ that contains the desired predicted elements for both tasks. In this work, we propose a hypothesis suggesting a correlation between sentiments associated with target objects. To support the model in distinguishing sentiment values for the two tasks, we introduce two binary auxiliary tasks. The objective of the first task is to identify whether the sentiment values of Task 1 and Task 2 align. Meanwhile, the second task aims to determine whether the sentiment values between the companies and consumers are equivalent. The output of both auxiliary tasks is formatted as either "True" or "False" labels. The text to text format is represented as follows:

$$headline \rightarrow target[e]p_{target}[e]p_{companies}[e]p_{consumers}[e]aux_1[e]aux_2 \tag{1}$$

In this paper, we utilize the pre-trained generative mT5 model [12] with encoder-decoder architecture for the following reasons: (1) First, the mT5 is the encoder-decoder architecture which can perform well with most of text generation tasks such as machine translation [13] and aspect-based sentiment analysis [6]. Second, the mT5 model is trained on a vast collection of natural text from 101 languages, sourced from the publicly accessible Common Crawl web scrape. Notably, Spanish (es) is one of the three languages with a substantial amount of training data in the variants of the mT5 model [12]. Third, representing the output of two tasks in this shared task as a natural language sentence becomes a straightforward process while fine-tuning the mT5 models. Given a new headline $\mathbf{x}$, the encoder component will map it to the sequence of contextual embedding $\boldsymbol{e}$. Then, the decoder assists to the output of the encoder layer to calculate the conditional probability distribution $p_\theta(\mathbf{y}|\mathbf{e})$ of the target sentence $\boldsymbol{y}$ where θ is the parameter weight. Finally, a softmax layer is applied to the output of the decoder to obtain the probability distribution for the next token y_i as:

$$p_\theta(y_i|e, y_{0:i-1}) = softmax(W^T y_i) \tag{2}$$

where W is a matrix to map the prediction y_i to a vector. In our work, we initialize the θ with the pre-trained parameter weights of mT5 models [12] and fine-tune the parameters to maximize the log-likelihood $p_\theta(\mathbf{y}|\mathbf{e})$.

4. Experimental Setup

4.1. Data and Evaluation Metrics

We only use the official training set [14], which is provided by the organizers, to train our models for two challenges in the shared-task. Table 1 presents the general statistics of the training and testing set, while Table 2 shows the statistical information towards the polarity classes for three sentiments tasks. As depicted in Table 2, it is evident that there exists an imbalance among the sentiment classes, along with distinct distribution variations between the two sets. These imbalances and distribution variations challenge participating teams when addressing sentiment-related problems within the datasets.

Table 1
The general information of official datasets in our experiments

Information	Training set	Testing set
Number of samples	6 359	1 621
Number of tokens	78 709	20 394
Number of vocabulary	14 740	6 041
Number of unique targets	3 357	1 068
The average length	12.38	12.58
The maximum length	35	36

Table 2
Statistics of polarity classes for three sentiment analysis tasks.

Tasks	Training set			Testing set		
	Positive	Neutral	Negative	Positive	Neutral	Negative
Target Sentiment	2815	606	2938	816	600	205
Companies Sentiment	646	3843	1870	523	822	276
Consumers Sentiment	897	4173	1289	553	803	265

4.2. System Settings

We implemented our framework based on the HuggingFace Transformer library [15]. All models were trained with 20 epochs. We use the learning rate of 1e-3 for the mT5-Small, mT5-based and

Table 3
Part of the official results of our final submission system with the Top 5 systems.

Ranking	Task 1: Financial Targeted Sentiment Analysis		Task 2: Financial SA for Companies & Consumers	
	Team	**F1-score**	**Team**	**F1-score**
Top 1	abc111	79.2244	lli-uam	64.2349
Top 2	lli-uam	79.2172	sjzafra	63.4901
Top 4	sjzafra	77.8002	abc111	57.5015
Top 5	AnkitSinghRaikuni	55.4211	fanchuyi	47.2685
Ours (Top 3)	**ABCD Team**	**78.2175**	**ABCD Team**	**61.0373**

1e-4 for mT5-large and mT5-XL. The batch size was set to {32,16} based on the size of pre-trained language models. The beam width was set to 5. We did not utilize any development data to tuning models. The AdamW optimizer was selected to optimize our models. The maximum input and output sequence lengths have been configured as 70 and 48 tokens, respectively. For our experiments, we set a fixed random seed of 42 to train the models. As the Spanish language is not familiar to us, we did not apply any pre-processing methods to the entire dataset except eliminating multiple spaces in the text.

5. Main results

The official results and the results of the top systems are shown in Table 3. Our best model achieves the Top 3 ranking in both challenges during the final round. Specifically, for Task 1 - identifying the target term and its corresponding sentiment, our model achieved a result of 78.2175% in terms of F1-score, which is lower than the F1 scores of the Top 1 and Top 2 teams, which are -1.0069% and -0.9997%, respectively. Regarding Task 2 - classifying the sentiment for the companies and consumers, our model also demonstrates the competition by achieving an F1 score of 61.0373%, which is lower than the top 1 team's model by -3.1976%. For the three sentiment analysis tasks, we achieved the F1-score of 78.9838% (Top 3), 58.8635% (Top 2) and 63.2111 (Top 3) for the target sentiment, companies sentiment and consumers sentiment tasks, respectively.

Table 4 shows the overall results of our submission model and other variants on the test set on two challenges and three sentiment analysis tasks. Overall, it can see that the performance of the model is improved when using language models of large size. As mentioned in the previous work [12], larger models are trained with a larger vocabulary in mT5 variants. Consequently, these larger models have the ability to capture and represent a wider range of linguistic patterns and contextual information for the given text input. However, this also poses a challenge regarding computational resources and model storage for large language models. As seen in Table 4, the models trained on both the original tasks and two auxiliary tasks exhibit better performance compared to models solely fine-tuned on the original tasks. This demonstrated our two auxiliary tasks improved the overall performance.

Figure 2 presents the confusion matrix for three sentiment analysis tasks in our final submis-

Table 4
Results from various mT5 variant models are presented for two challenges and four tasks.

Model	F1 Task 1	F1 Task 2	Target	F1 Target Sentiment	F1 Companies Sentiment	F1 Consumers Sentiment
Models fine-tuned on original tasks						
mT5-Small	66.8934	42.1374	71.7443	61.4424	36.5391	47.7357
mT5-Base	67.9454	50.9910	74.2953	61.5954	47.9879	53.9941
mT5-Large	69.5612	52.5217	76.8780	62.2444	47.7485	57.2949
mT5-XL	77.2777	60.7960	84.7134	69.8420	60.0068	61.5851
Models fine-tuned on original tasks and two auxiliary tasks						
mT5-Small	66.7274	43.3627	71.8191	61.6357	37.1061	49.6192
mT5-Base	68.0629	51.2467	74.4745	61.6512	48.3223	54.1711
mT5-Large	70.8644	54.8205	74.88919	66.8368	51.7291	57.9118
mT5-XL	78.2175	61.0373	85.4511	70.9838	58.8635	63.2111

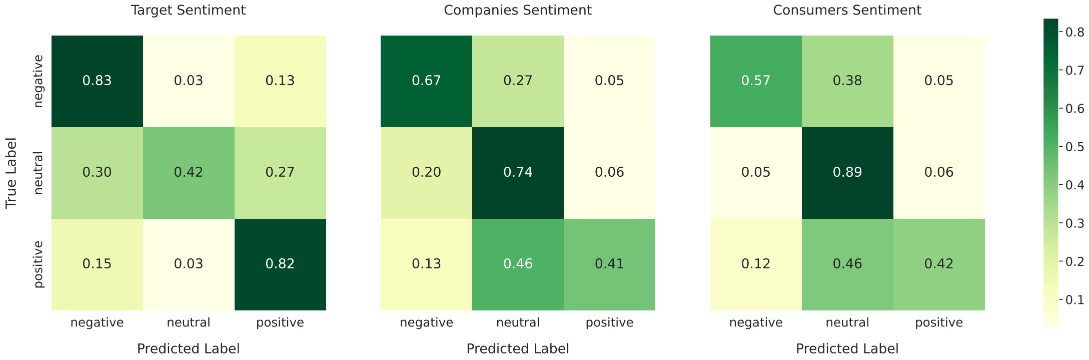

Figure 2: The confusion matrix of three sentiment analysis tasks on our official submission.

sion. In the sentiment analysis task for target terms, our model performs well in classifying "positive" and "negative" sentiments but struggles with the "neutral" class. The rate of incorrectly predicting neutral labels as positive versus negative labels is relatively high, with a ratio of approximately 30% and 27%, respectively. This result contrasts with the findings from Task 2, which focused on sentiment analysis for companies and consumers. As shown in Figure 2, our approach achieves the highest accuracy on classifying neutral labels, while its performance on positive and negative labels is comparatively lower. A significant percentage of misclassified positive and negative labels, particularly in the "positive" class, are incorrectly predicted as neutral labels in both tasks. One of the reasons for this result could be the imbalance between polarity classes, as shown in Table 2, within the training set.

6. Conclusion and Future Work

In this paper, we describe our submission system for the IBERLEF 2023 Task - FinancES: Financial Targeted Sentiment Analysis, which achieved the Top 3 in both sub-tasks in the shared task.

Instead of solving two challenges through classification approaches, we present a generative approach based on the contextual pre-trained language model to solve tasks at the same time. Besides, we design two simple auxiliary tasks to provide more information between two tasks simultaneously. Based on our experimental results and analysis, we believe this approach can be widely applied to other domains for targeted sentiment analysis. Due to the limitations in computational resources, we cannot explore large language models such as mT5-XXL [12] or BLOOM [16] models. However, we believe that fine-tuning these models can improve prediction effectiveness. Additionally, applying effective text pre-processing techniques can also enhance performance on the test set.

Acknowledgments

Dang Van Thin was funded by the Master, PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2022.TS120.

References

[1] S. M. Jiménez-Zafra, F. Rangel, M. Montes-y Gómez, Overview of IberLEF 2023: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org, 2023.

[2] J. A. García-Díaz, Almela, F. García-Sánchez, G. Alcaráz Mármol, M. J. Marín-Pérez, R. Valencia-García, Overview of FinancES 2023: Financial Targeted Sentiment Analysis in Spanish, Procesamiento del Lenguaje Natural 71 (2023).

[3] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, The Journal of Machine Learning Research 21 (2020) 5485–5551.

[4] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 7871–7880. URL: https://aclanthology.org/2020.acl-main.703. doi:10.18653/v1/2020.acl-main.703.

[5] H. Yan, T. Gui, J. Dai, Q. Guo, Z. Zhang, X. Qiu, A unified generative framework for various NER subtasks, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online, 2021, pp. 5808–5822. URL: https://aclanthology.org/2021.acl-long.451. doi:10.18653/v1/2021.acl-long.451.

[6] H. Yan, J. Dai, T. Ji, X. Qiu, Z. Zhang, A unified generative framework for aspect-based sentiment analysis, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on

Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online, 2021, pp. 2416–2429. URL: https://aclanthology.org/2021.acl-long.188. doi:10.18653/v1/2021.acl-long.188.

[7] S. Yıldırım, D. Jothimani, C. Kavaklioğlu, A. Başar, Deep learning approaches for sentiment analysis on financial microblog dataset, in: 2019 IEEE International Conference on Big Data (Big Data), IEEE, 2019, pp. 5581–5584.

[8] C.-C. Lee, Z. Gao, C.-L. Tsai, Bert-based stock market sentiment analysis, in: 2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), 2020, pp. 1–2. doi:10.1109/ICCE-Taiwan49838.2020.9258102.

[9] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev, D. Trajanov, Evaluation of sentiment analysis in finance: From lexicons to transformers, IEEE Access 8 (2020) 131662–131682. doi:10.1109/ACCESS.2020.3009626.

[10] K. L. Kohsasih, B. H. Hayadi, Robet, C. Juliandy, O. Pribadi, Andi, Sentiment analysis for financial news using rnn-lstm network, in: 2022 4th International Conference on Cybernetics and Intelligent System (ICORIS), 2022, pp. 1–6. doi:10.1109/ICORIS56080.2022.10031595.

[11] K. Ong, W. van der Heever, R. Satapathy, G. Mengaldo, E. Cambria, Finxabsa: Explainable finance through aspect-based sentiment analysis, arXiv preprint arXiv:2303.02563 (2023).

[12] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, C. Raffel, mT5: A massively multilingual pre-trained text-to-text transformer, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online, 2021, pp. 483–498. URL: https://aclanthology.org/2021.naacl-main.41. doi:10.18653/v1/2021.naacl-main.41.

[13] O. Agarwal, M. Kale, H. Ge, S. Shakeri, R. Al-Rfou, Machine translation aided bilingual data-to-text generation and semantic parsing, in: Proceedings of the 3rd International Workshop on Natural Language Generation from the Semantic Web (WebNLG+), Association for Computational Linguistics, Dublin, Ireland (Virtual), 2020, pp. 125–130. URL: https://aclanthology.org/2020.webnlg-1.13.

[14] P. Ronghao, J. A. García-Díaz, F. García-Sánchez, R. Valencia-García, Evaluation of transformer models for financial targeted sentiment analysis in spanish, PeerJ Computer Science 9 (2023) e1377. URL: https://doi.org/10.7717/peerj-cs.1377. doi:10.7717/peerj-cs.1377.

[15] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, A. Rush, Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 38–45. URL: https://aclanthology.org/2020.emnlp-demos.6. doi:10.18653/v1/2020.emnlp-demos.6.

[16] T. L. Scao, A. Fan, C. Akiki, E. Pavlick, S. Ilić, D. Hesslow, R. Castagné, A. S. Luccioni, F. Yvon, M. Gallé, et al., Bloom: A 176b-parameter open-access multilingual language model, arXiv preprint arXiv:2211.05100 (2022).