ISTC-CNR at EmotivITA: Towards Better Dimensional and multi-dimensional Analysis of VAD Emotions*

Luana Bulla[1,2,*,†], Misael Mongiovì[2,†]

[1]University of Catania, Italy
[2]National Research Council, Italy

Abstract
Emotions are an integral part of human communication, and accurately detecting and interpreting them from textual data holds significant potential for numerous applications. The Valence-Arousal-Dominance (VAD) dimensional model provides a rich framework for capturing the nuanced emotional states conveyed through text. This paper presents an in-depth exploration of text-based emotion detection using VAD analysis and machine learning techniques. In this paper, we propose a novel machine-learning model specifically designed to detect VAD dimensions from textual data. Through empirical evaluation and comparison with existing methods, we analyze the results of our proposed model in accurately identifying emotions expressed in text. This paper describes the ISTC-CNR participation in the EmotivITA task at EVALITA 2023, showcasing our team's efforts and findings in the context of the competition.

Keywords
NLP, Machine Learning, NLI, Emotion Regression

1. Introduction

Emotions constitute a complex facet of human experience, influencing our perception, behavior, and overall well-being. Traditionally, emotions have been studied using categorical models [1], which assign discrete labels to distinct emotional states. However, the limitations of such models in capturing the richness and variability of emotions led to the development of dimensional models. Among these, the VAD dimensional model [2] has gained considerable attention due to its ability to provide a multi-dimensional representation of emotions. According to the VAD paradigm, emotions may be classified along three main dimensions: valence, which describes the pleasantness or unpleasantness of an emotional experience; arousal, which reflects the level of physiological activation associated with emotion; and dominance, which denotes the degree of control or influence exerted by an emotion. This dimensional framework offers a more fine-grained understanding of emotional experiences, enabling researchers to capture the subtle nuances and interplay of emotions.

The primary objective of this paper is to explore text-based emotion detection by leveraging the VAD dimensional model through the fine-tuning of a multi-lingual MNLI-XML-RoBERTa model. The MNLI-XML-RoBERTa model is based on the well-known architecture of XML-RoBERTa [3] and has been fine-tuned on an Italian version of the MNLI dataset [4] for solving a Natural Language Inference (NLI) task. In this study, we investigate the model's inference capability and its adaptability to handle multiple languages to address the VAD regression task in both single- and multi-dimensional scenarios. To achieve our goal, we apply the model to the Italian version of the EmoBank dataset (EmoITA) [5], which consists of 10,000 short text documents tagged with VAD emotional scores. We propose a specific configuration of the MNLI-XML-RoBERTa model for both mono- and multi-dimensional scenarios. In the first scenario, we train three separate models, each capable of predicting one of the Valence, Arousal, or Dominance values using only the target dimension during training. For multi-dimensional emotional analysis, we develop a single model that predicts all three values jointly by utilizing training data from all three dimensions.

We present our model at EmotivITA[1], a task in the EVALITA 2023 campaign [6], validating and discussing our results while comparing them with the baseline provided by the organizers for both sub-task A (Dimensional emotion regression) and B (Multi-dimensional emotion regression)[2].

The paper is organized as follows: Section 2 introduces our novel machine learning model for VAD-based emotion detection, providing an insightful description of the model architecture. Section 3 presents the experimental setup, evaluation metrics, and results, comparing the

EVALITA 2023: 8th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian, Sep 7 – 8, Parma, IT
*Corresponding author.
†These authors contributed equally.
luana.bulla@phd.unict.it (L. Bulla); misael.mongiovi@cnr.it (M. Mongiovì)
0000-0003-1165-853X (L. Bulla); 0000-0003-0528-5490 (M. Mongiovì)

[1]https://sites.google.com/view/emotivita
[2]https://sites.google.com/view/emotivita/the-tasks

performance of the VAD MNLI-XML-ROBERTA-based system with the baseline BERT-based model provided by the organizers. Finally, Section 4 summarizes the key findings, discusses the practical implications, and outlines potential directions for future research.

2. Description of the Systems

We describe our NLI-based method for recognizing fine-grained emotions along the continuous VAD dimensions. Our approach encompasses both the dimensional (sub-task A) and multi-dimensional (sub-task B) VAD regression tasks, following a unified methodology framework. To accomplish this, we introduce a supervised model developed within a constrained mode, leveraging the EmoITA. Section 2.1 provides a comprehensive overview of our methodology, offering detailed insights into the key components and techniques employed in our NLI-based approach. Section 2.2 provide an overview of the prompts utilized during the training phase.

2.1. VAD MNLI-XML-RoBERTa model

We present the VAD MNLI-XML-RoBERTa model, a fine-tuned version of the multilingual XML-RoBERTa system, which has been developed using an automatically translated Italian version of the MNLI corpus[3]. To adapt the model for the regression task, we replace the last linear layer of the MNLI-XML-RoBERTa model. During training, the parameters of the last linear layer are initialized from scratch and learned, while the remaining parameters are fine-tuned. To tune the hyperparameter, we train the model on a portion of the training set provided by the organizers in a preliminary phase. Our model utilizes input sentences as premises, accompanied by predefined fixed prompts that serve as hypotheses. These prompts aim to guide the model's training process by focusing on specific assignments. The VAD MNLI-XML-RoBERTa model is trained with a learning rate of 1e-5, a batch size of 64, and AdamW as the optimizer. The Mean Squared Error (MSE) is adopted as the loss function, measuring the disparity between the model's predictions and the actual values[4].

2.2. Prompting

We introduce a set of tailored prompts designed to predict the values of valence, arousal, and dominance in text. For the dimensional emotion analysis task, we develop three prompts, each targeting a specific emotion dimension. The prompt for valence is "quanta positività esprime la frase?" (how much positivity does the sentence convey?). For arousal, the prompt is "quanto è eccitante la frase?" (how exciting is the sentence?). Lastly, for dominance, the prompt is "quanto è controllata l'emozione?" (how controlled is the emotion?). These prompts are crafted to succinctly encapsulate the essence of each emotional dimension using easily understood language. Their purpose is to guide the model's training process toward discerning specific emotional dimensions accurately. In the multi-dimensional emotion analysis task, we utilize a single prompt to predict all emotion dimensions. The prompt used is "valence, arousal, dominance dell'emozione?" (valence, arousal, dominance of the emotion?). In this case, the prompt assists the model in comprehending the comprehensive framework of the task during the training phase, providing contextual information. By leveraging these prompts, our objective is to enhance the model's capacity to capture and comprehend the intricate nuances of emotional dimensions in text.

3. Results and Evaluation

We conduct an experimental analysis to assess the effectiveness of our proposed approach and compare it with the baseline method provided by the competition's organizers using EmoITA [5]. In Section 3.1, we present the results of the dimensional emotion analysis, focusing on the individual evaluation of valence, arousal, and dominance. Here, we examine the performance of our approach and provide a comprehensive analysis of its efficacy. Moving on to Section 3.2, we shift our focus to multi-dimensional emotion analysis. In this section, we evaluate the performance of our approach in capturing and predicting emotional dimensions collectively.

3.1. Dimensional Emotion Analysis

We train three separate VAD MNLI-XML-RoBERTa models, each focused on detecting the degree of valence, arousal, and dominance values individually. During training, each model focuses on a specific target dimension and is associated with a specific prompt (Sect. 2).

Tables 3.1 and 3.1 showcase the results obtained by evaluating our MNLI-based model on 2′063 items from the Italian EmoBank test set provided by the organizers. For the dimensional emotion analysis task, we submitted two different runs corresponding to distinct configurations of the fine-tuned VAD MNLI-XML-RoBERTa model. During the model training phase, the first run incorporates 99% of the data from the development set, while the second run utilizes the entire development set. Each emotional dimension is evaluated based on Pearson's correlation coefficient (Table 3.1) and Mean Absolute Errors (Table 3.1). To assess the effectiveness of our model, we

[3]https://huggingface.co/Jiva/xlm-roberta-large-it-mnli
[4]https://pytorch.org/docs/stable/generated/torch.nn.MSELoss.html

Table 1
Pearson's r correlation score for valence, arousal and dominance values for the baseline BERT-based model and the MNLI-based systems in the Dimensional emotional detection task.

Model	Valence	Arousal	Dominance
Baseline	0.807	**0.643**	**0.643**
VAD MNLI-XML-RoBERTa (run 1)	0.800	0.593	0.609
VAD MNLI-XML-RoBERTa (run 2)	**0.809**	0.587	0.624

Table 2
Mean Absolute Error (MAE) score for valence, arousal and dominance values for the baseline BERT-based model and the MNLI-based systems in the Dimensional emotional detection task.

Model	Valence	Arousal	Dominance
Baseline	0.313	0.321	0.285
VAD MNLI-XML-RoBERTa (run 1)	0.393	0.432	0.399
VAD MNLI-XML-RoBERTa (run 2)	0.382	0.433	0.387

compare its performance against a baseline BERT-based model provided by the organizers. Our NLI-based model demonstrates slightly higher Pearson's correlation values for valence compared to the baseline, while the other metrics show slightly lower scores. Notably, valence exhibits the highest performance in terms of correlation, while arousal and dominance present more challenges, with the baseline model achieving a correlation of only 0.64%.

3.2. Multi-dimensional Emotion Analysis

We present the performance of models designed for detecting valence, arousal, and dominance in a multi-dimensional setting. In this setting, both the baseline and MNLI-based models predict all three values by utilizing all three VAD dimensions during the training phase. For this task, we submitted a single run where we employed all the available development set data for training the model. We report the performances of the VAD MNLI-XML-RoBERTa system for each emotion dimension, evaluating them in terms of Pearson's correlation coefficient (Table 3.2) and Mean Absolute Error (Table 3.2). We compare our results with the baseline performance provided by the organizers.

Both the baseline and NLI-based models demonstrate superior performance in detecting valence, achieving a correlation of approximately 80%. As previously mentioned, the scores for the arousal and dominance dimensions are comparatively lower than valence. Specifically, the highest correlation values observed for arousal and dominance values are 0.652 and 0.654, respectively.

4. Discussion

As anticipated, performance varied significantly across different emotion values in both dimensional (i.e. sub-task A) and multi-dimensional (i.e. sub-task B) emotion regression tasks. The concept of valence demonstrated consistent interpretation, resulting in higher correlation scores for all models in both analyses. Conversely, detecting concepts like "arousal" and "dominance" proved more challenging in the context of short text documents. Comparing the two models, the BERT-based model yielded the best results, while the NLI-based model exhibited slightly lower performance in both tasks. This discrepancy can be attributed to the multilingual nature of the MNLI-XML-RoBERTa model, which grants it greater adaptability but limits its ability to identify nuanced and ambiguous terms like "dominance" and "excitement" within text. Additionally, the NLI-based models employed a generic prompt for every item in the dataset as a hypothesis. This approach potentially underutilized the model's capabilities as the prompt lacked the necessary explanatory power to effectively tackle challenging regression tasks. In conclusion, while the models demonstrated strengths in capturing valence, improvements are needed to enhance their performance in discerning arousal and dominance. This includes refining the model's adaptability to diverse linguistic contexts and developing more informative prompts for comprehensive analysis of emotional dimensions.

5. Conclusion

In our study, we propose a novel approach for detecting valence, arousal, and dominance values in natural language short text using an MNLI-XML-RoBERTa-based model. Our method encompasses both dimensional and

Table 3
Pearson's r correlation score for valence, arousal and dominance values for the baseline BERT-based model and the MNLI-based systems in the multi-dimensional emotional detection task.

Model	Valence	Arousal	Dominance
Baseline	0.811	**0.652**	**0.654**
VAD MNLI-XML-RoBERTa	0.800	0.594	0.623

Table 4
Mean Absolute Error (MAE) score for valence, arousal and dominance values for the baseline BERT-based model and the MNLI-based systems in the multi-dimensional emotional detection task.

Model	Valence	Arousal	Dominance
Baseline	0.859	0.859	0.859
VAD MNLI-XML-RoBERTa	0.400	0.421	0.367

multi-dimensional regression scenarios, leveraging a fine-tuned XML-RoBERTa classifier trained and evaluated on the EmoITA. We present the results of both the dimensional and multi-dimensional configurations' models and compare them with the performance of the BERT-based model provided by the competition's organizers. This evaluation provides valuable insights into the effectiveness of our proposed approach in capturing and analyzing emotional dimensions in textual data.

References

[1] F. A. Acheampong, C. Wenyu, H. Nunoo-Mensah, Text-based emotion detection: Advances, challenges, and opportunities, Engineering Reports 2 (2020) e12189.

[2] J. A. Russell, A. Mehrabian, Evidence for a three-factor theory of emotions, Journal of research in Personality 11 (1977) 273–294.

[3] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, CoRR abs/1911.02116 (2019). URL: http://arxiv.org/abs/1911.02116. arXiv:1911.02116.

[4] A. Williams, N. Nangia, S. Bowman, A broad-coverage challenge corpus for sentence understanding through inference, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 1112–1122. URL: https://aclanthology.org/N18-1101. doi:10.18653/v1/N18-1101.

[5] G. Gafà, F. Cutugno, M. Venuti, EmotivITA at EVALITA2023: Overview of the Dimensional and Multidimensional Emotion Analysis Task, in: M. Lai, S. Menini, M. Polignano, V. Russo, R. Sprugnoli, G. Venturi (Eds.), Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), CEUR.org, Parma, Italy, 2023.

[6] M. Lai, S. Menini, M. Polignano, V. Russo, R. Sprugnoli, G. Venturi, Evalita 2023: Overview of the 8th evaluation campaign of natural language processing and speech tools for italian, in: Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), CEUR.org, Parma, Italy, 2023.