EVALITA 2023

EmotivITA: Dimensional and Multi-dimensional Emotion Analysis in Italian Texts

Task Guidelines

Giovanni Gafà[1], Francesco Cutugno[2], Marco Venuti[1]

[1]University of Catania, Catania
[2]University of Naples Federico II, Naples

Contents

1 Introduction

In the past few years, affective computing has become a prominent area in Natural Language Processing. While early studies focused on predicting the semantic polarity of a text, over the years research moved towards more fine-grained modelling of affective states. To make just a few examples in EVALITA recent history: aspect-based sentiment analysis, which aims to detect polarity classes in relation to multiple aspects of an opinion item (Basile et al., 2018); automatic classification of specific communicative intentions, such as irony and sarcasm (Cignarella et al., 2018); stance detection, expected to determine the orientation of a writer in favor or against certain topics of interest (Cignarella et al., 2020).

Dimensional emotion analysis, however, has never been addressed in EVALITA campaigns. Dimensional emotion analysis is based on dimensional models of emotions rather than categorical ones. While the latter describes affective states with limited sets of discrete emotions, such as that defined by Ekman (Ekman, 2005), the former makes use of a continuous scale spanning through three independent dimensions: Valence (degree of pleasure), Arousal (degree of excitement) and Dominance (degree of control over the situation) - VAD from now on.

The VAD model has received increasing attention in NLP tasks in English and has been used both for dimensional analysis (Park et al., 2021; Wang et al., 2016), aiming at predicting one dimension at a time, and multi-dimensional analysis (Mukherjee et al., 2021; Zhu et al., 2019), which tries to determine all three dimensions at the same time. Emotion regression has been also addressed in international evaluation campaigns, such as SemEval 2018 (Mohammad et al., 2018). Nevertheless, while a limited number of Italian corpora annotated with categorical models are available, to the best of our knowledge, no VAD annotated dataset exists for the Italian language.

We fill that gap with EMOITA: the first *corpus* of Italian sentences annotated with their VAD values. EMOITA is the Italian version of EMOBANK (Buechel & Hahn, 2017), the largest English VAD dataset to date. Sentences from EMOBANK have been translated in Italian and revised by a group of Italian researchers affiliated to Urban/Eco (a research center associated with the University of Naples Federico II) and the Humanities Department of the University of Catania. Researchers also annotated each sentence with new VAD values.

EmotivITA tasks aim to exploit the EMOITA dataset and promote dimensional emotion analysis of Italian texts.

2   Task Description

EmotivITA includes two tasks, both constraint and unconstraint:

- **Subtask A:** Dimensional emotion regression: Prediction of Valence, Arousal and Dominance values based on a set of Italian sentences and annotations, using only the target annotated dimension for training - so, for instance, when predicting Valence participant systems may only use Valence values for training; the same holds for Arousal and Dominance.
- **Subtask B:** Multi-dimensional emotion regression: Prediction of Valence, Arousal and Dominance values based on a set of Italian sentences and annotations, using all mentioned dimensions for training - so participant systems should determine Valence, Arousal and Dominance values at once, using Valence, Arousal and Dominance values for training. Subtask B intend to study and exploit potential correlations between Valence, Arousal and Dominance, which have been discussed in literature (Kuppens et al., 2013; Warriner et al., 2013).

Both subtasks are regression problems, so participating teams should provide in the output the sentence *id* and three real numbers between 1 and 5, relative to the three predicted dimensions.

3   The dataset

To realize the EMOITA dataset, the English EMOBANK *corpus* has been translated and annotated by native Italian speakers, with seven different annotations for each sentence and each of the three dimensions. The development set include 8,000 sentences together with the average values from the 7 annotations. Valence, Arousal, and Dominance vary continuously from a minimum of 1 to a maximum of 5, with two decimal places.

The development set is provided as a CSV comma-separated file, reporting the following fields:

<div align="center">id, V, A, D</div>

where:

- id denotes the unique identifier of the sentence.
- V denotes the average Valence value annotated for the sentence.
- A denotes the average Arousal value annotated for the sentence.
- D denotes the average Dominance value annotated for the sentence.

As follows, we provide an example of a subset of training data instances:

id	text	V	A	D
260	Jet si capovolge durante una tormenta, nessun morto	3.33	4.17	3
261	Certo, risposi.	3.83	3.17	3.83
262	Youngstown fu probabilmente colpita più duramente di qualsiasi altra città dal collasso dell'industria siderurgica, ma non era sola.	2	4.25	2.42
263	Abbigliamento prêt-à-porter.	4	3.42	3.67

4   Submission format

The participants should provide an output file in CSV comma-separated format reporting:

- the team's name.
- the addressed Subtask (A or B).
- the ID of the run (1 or 2).
- and the type of the run (constrained or unconstrained).

For instance, if a team's first run addresses the Subtask A in constrained modality, the name of the file should be:

TeamName_SubtaskA_constrained_run1.csv

The submission files for both subtasks should contain 4 comma-separated fields:

<div align="center">id, V, A, D</div>

Each line must then contain four comma-separated values: the id of the sentence identifier number listed in the test set, followed by the predicted Valence, Arousal, and Dominance values for that sentence; when reporting V, A, and D the dot should be used to indicate decimal places. Only two digits after the decimal point will be considered, so the values should be rounded (off or up) accordingly: e.g., 4.3449 to 4.34, 4.3450 to 4.35. Below, we report a toy example of a submitted run.

id	V	A	D
1	3.13	2.17	3.00
2	2.83	3.10	4.83

The ZIP file containing the relevant runs should be submitted during the evaluation window (12th – 19th May 2023) to the following email address:

giovanni.gafa@gmail.com

with the subject:

"Submission: EmotivITA - EVALITA2023 - TeamName"

The name of the zip file should have the form: TeamName.zip

It is important to notice that:

- Participation is allowed to all two subtasks (Subtask A and Subtask B) or only to one of them. Participation to Subtask A is anyway strongly recommended, in order to have a common basis for comparison.
- Each participating team will be allowed to submit a maximum of 2 runs for each subtask.
- All runs can be produced according to the constrained modality or the unconstrained modality; we ask anyway to specify the type of the run. In constrained modality, the only annotated data to be used for training and tuning the systems are those distributed by the organizers. Other linguistic resources, e.g., word embeddings, and lexicons are instead allowed. In unconstrained modality, annotated external data can be also employed. All external resources should be described in the systems' reports.

Also note that registration is non-binding but participation implies:

- Submitting your predictions during the evaluation window
- Writing a report (paper of max 6 pages) that summarizes your proposed method(s), the data resource(s) used, details on the experiments, and an analysis of the results. The report has to be sent to task organizers by 14th June 2023 and, if accepted, it should be officially published in the EVALITA workshop proceedings.

5 Evaluation

During the evaluation window (12th – 19th May 2023), the test data containing the 2062 remaining sentences from the EMOITA *corpus* will be released, without the VAD values. For each subtask, a separate official ranking will be provided and the evaluation will be performed using the established measures for such tasks, namely, Pearson's r and Mean Absolute Error. Ranking of the systems will be based mainly on Pearson's r, but Mean Absolute Error will be reported and we consider to weight rankings on its basis.

Runs produced in constrained and unconstrained modality will all be evaluated together.

Finally, the baselines' systems for both subtasks have been built training to regression a BERT model available on HuggingFace, with a learning rate of 1e-05.

6 Final remarks

Datasets and baselines will be made available on EmotivITA Google Group: https://groups.google.com/g/emotivita-evalita2023/

If you have any question or problem, please open a thread on the Google Group mailing list, or you can contact Giovanni Gafà: giovanni.gafa@gmail.com.

References

Basile, P., Croce, D., Basile, V., & Polignano, M. (2018). Overview of the EVALITA 2018 Aspect-based Sentiment Analysis task (ABSITA). In T. Caselli, N. Novielli, V. Patti, & P. Rosso (Eds.), *EVALITA Evaluation of NLP and Speech Tools for Italian* (pp. 10–16). Accademia University Press. https://doi.org/10.4000/books.aaccademia.4451

Buechel, S., & Hahn, U. (2017). EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis. In M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 2: Short Papers* (pp. 578–585). Association for Computational Linguistics. https://doi.org/10.18653/v1/E17-2092

Cignarella, A. T., Frenda, S., Basile, V., Bosco, C., Patti, V., & Rosso, P. (2018). Overview of the EVALITA 2018 Task on Irony Detection in Italian Tweets (IronITA). In T. Caselli, N. Novielli, V. Patti, & P. Rosso (Eds.), *EVALITA Evaluation of NLP and Speech Tools for Italian* (pp. 26–34). Accademia University Press. https://doi.org/10.4000/books.aaccademia.4470

Cignarella, A. T., Lai, M., Bosco, C., Patti, V., & Rosso, P. (2020). SardiStance @ EVALITA2020: Overview of the Task on Stance Detection in Italian Tweets. *EVALITA Evaluation of NLP and Speech Tools for Italian - December 17th, 2020*. https://doi.org/10.4000/books.aaccademia.7084

Ekman, P. (2005). Basic Emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 45–60). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470013494.ch3

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, *139*(4), 917–940. https://doi.org/10.1037/a0030811

Mohammad, S., Bravo-Marquez, F., Salameh, M., & Kiritchenko, S. (2018). SemEval-2018 Task 1: Affect in Tweets. *Proceedings of The 12th International Workshop on Semantic Evaluation*, 1–17. https://doi.org/10.18653/v1/S18-1001

Mukherjee, R., Naik, A., Poddar, S., Dasgupta, S., & Ganguly, N. (2021). Understanding the Role of Affect Dimensions in Detecting Emotions from Tweets: A Multi-task Approach. In F. Diaz 0001, C. Shah, T. Suel, P. Castells, R. Jones, & T. Sakai (Eds.), *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021* (pp. 2303–2307). ACM. https://doi.org/10.1145/3404835.3463080

Park, S., Kim, J., Ye, S., Jeon, J., Park, H., & Oh, A. (2021). Dimensional Emotion Detection from Categorical Emotion. In M.-F. Moens, X. Huang, L. Specia, & S. W. Yih (Eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021* (pp. 4367–4380). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.358

Wang, J., Yu, L.-C., Lai, K. R., & Zhang, X. (2016). Dimensional Sentiment Analysis Using a Regional CNN-LSTM Model. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 225–230. https://doi.org/10.18653/v1/P16-2037

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Zhu, S., Li, S., & Zhou, G. (2019). Adversarial Attention Modeling for Multi-dimensional Emotion Regression. In A. Korhonen, D. R. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers* (pp. 471–480). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1045