

Abhinav Patil, Sam Briggs, Tara Wueger, D. D. O’Connell

Department of Linguistics, University of Washington
Seattle, WA

{abhinavp, briggs3, taraw28, danieloc}@uw.edu

Abstract

TODO

1 Introduction

2 Task Description

We will participate in the ”Multimodal Abusive Language Detection and Sentiment Analysis: DravidianLangTech@RANLP 2023” shared task hosted on CodaLab (B. et al., 2023).¹

This shared task has two subtasks: abusive language detection in Tamil, and sentiment analysis in both Tamil and Malayalam (members of the Dravidian language family). Both tasks are multimodal, consisting of videos paired with files containing just their audio tracks as well as (sometimes partial) text transcripts. For the purposes of this class, we will be working on the second subtask of sentiment analysis in Tamil and Malayalam.

Sentiment analysis consists of categorizing a unit of language about a certain topic according to the author’s attitude towards that topic. This task can be constructed in many ways, e.g. as a binary classification task (i.e., categorizing language in two classes, Positive and Negative sentiment), as a multi-class classification problem, or, as in our case, an ordinal regression problem. Sentiment analysis has a rich history and many methods have been used (Cui et al., 2023).

2.1 Primary Task: Sentiment Analysis of Text Data

Our task is to categorize text written in either Tamil or Malayalam into 5 different categories: Highly Negative, Negative, Neutral, Positive, or Highly Positive. The data is ordinal, that is, the categories are discrete and there exists a total order

over them, but the distances between the categories are taken to be unknown or not well-defined. Thus, we formulate the task as one of ordinal regression, where the dependent variable can take one of the aforementioned values.

Previous work on sentiment analysis on Dravidian languages has been done using code-switched data (English and either Tamil or Malayalam) (Chakravarthi et al., 2021). Our data is not code-switched, and only contains one language. Like many tasks in Natural Language Processing, different Neural Network architectures have been used to perform sentiment analysis (Habimana et al., 2019).

2.2 Adaptation Task: Sentiment Analysis of Multimodal Data: Text, Audio, and Video

For our adaptation task, we will continue to work on sentiment analysis; however, we will be including audio and video data in addition to the text data from the primary task. This task will also be multilingual, using data from both Tamil and Malayalam. The same five categories will be used in the adaptation task as in the primary task.

Work on analyzing and indexing videos has been summarized in Snoek and Worring (2005). Additionally, Dimitri (2022) contains an extensive overview of multimodal integration methods and their applications, including for the purpose of sentiment analysis.

2.3 Data set

2.3.1 Description

The shared task organizers have provided, in total, 52 Tamil samples and 70 Malayalam samples. Each sample consists of a video containing speech in the given language, the corresponding audio file, and a (sometimes partial) transcript of the speech in the audio. From an initial analysis of a subset of the data samples, with the help of Google Translate,

¹The official citation has yet to be released as of April 7, 2023.

we tentatively believe the samples are all movie reviews, though this has not been stated by the task organizers explicitly.

The shared task organizers permit the use of other data sources and pretrained models as long as they are named and cited properly. We are still exploring the possibility of using other data; at the present time, it is unlikely we will do so. We are, however, likely to use pretrained models in some way.

2.3.2 Split

The task organizers have provided a train and dev split over the data set, which is further subdivided by language (Tamil and Malayalam). The 52 Tamil samples are split 42/10 train/dev while the Malayalam samples are split 60/10.

We believe that test (evaluation) data will be released at some point in the future, but as we are unsure of the date, for the time being we will be using the official dev data set as our evaluation data set, while holding out 20% portion of the official train data set as our dev data set. This gives us a (provisional) split of 34/8/10 and 50/10/10 for Tamil and Malayalam, respectively.

The official data can be found in Google Drive folder linked below:

- [Tamil train data](#)
- [Tamil dev data](#)
- [Malayalam train data](#)
- [Malayalam dev data](#)

2.4 Evaluation

The organizers of the competition have stated that they will use an F1 metric for evaluation purposes. They have not release evaluation tools or scripts at this time, but suggest teams use Sklearn’s classification report function (which lists precision, recall, F1, and a confusion matrix) in evaluating performance, suggesting that they may use the same function in evaluation tools (whether eventually released or not). We intend to follow their advice in this matter.

We note that F1 is a problematic metric for ordinal data, as it is ambivalent to the ordering of the class labels; for example, given a review whose true label is “highly positive,” an F1 score would equally penalize a model classifying it as “positive” as “highly negative.” However, since this is the

metric of choice for the task organizers, we will use it as our baseline evaluation metric (over the official “dev” data, which is our test data). Nevertheless, we are currently exploring other metrics we can use, both for evaluation of our end-to-end system, and for use as a loss metric when training our model(s), which would more accurately capture the inter-class ordering of our labels.

3 System Overview

4 Approach

5 Results

6 Discussion

7 Ethical Considerations

8 Conclusion

References

- Premjith B., Sowmya V., Joyithish Lal. G., Bharathi Raja Chakravarthi, K. Nandhini, Rajeswari Natarajan, Abirami Murugappan, and Bharathi B. 2023. [Multimodal Abusive Language Detection and Sentiment Analysis:DravidianLangTech@RANLP 2023](#).
- Bharathi Raja Chakravarthi, Ruba Priyadharshini, Vigneshwaran Muralidaran, Shardul Suryawanshi, Navya Jose, Elizabeth Sherly, and John P. McCrae. 2021. [Overview of the track on sentiment analysis for dravidian languages in code-mixed text](#). In *Proceedings of the 12th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE ’20*, page 21–24, New York, NY, USA. Association for Computing Machinery.
- Jingfeng Cui, Zhaoxia Wang, Seng-Beng Ho, and Erik Cambria. 2023. [Survey on sentiment analysis: evolution of research methods and topics](#). *Artificial Intelligence Review*.
- Giovanna Maria Dimitri. 2022. [A short survey on deep learning for multimodal integration: Applications, future perspectives and challenges](#). *Computers*, 11(11).
- Olivier Habimana, Yuhua Li, Ruixuan Li, Xiwu Gu, and Ge Yu. 2019. [Sentiment analysis using deep learning approaches: an overview](#). *Science China Information Sciences*, 63(1):111102.
- Cees GM Snoek and Marcel Worring. 2005. [Multimodal video indexing: A review of the state-of-the-art](#). *Multimedia tools and applications*, 25:5–35.

A Appendices