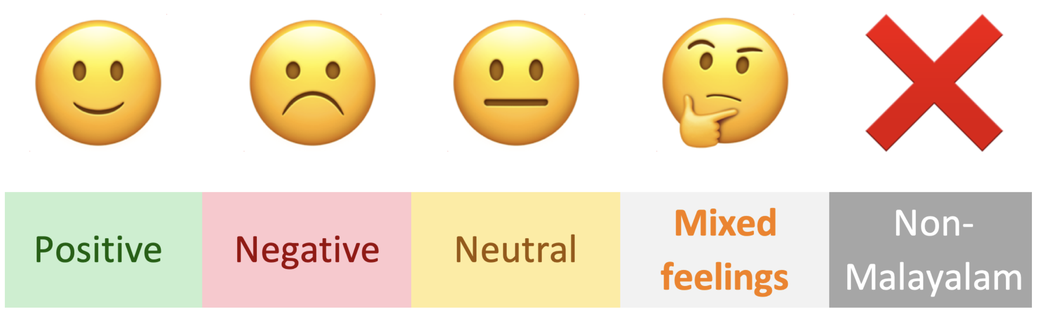
Exploring Reproducibility of Human-Labelled Data for Code-Mixed Sentiment Analysis

Sachin Sasidharan Nair, Tanvi Dinkar, Gavin Abercrombie

Introduction

- Re-implement a **human data collection study**.
- Task: **sentiment analysis** by using **code-mixed Malayalam-English** language data.
- **Identify** issues faced while reproducing the processes followed by the original study.
- **Verify** whether similar results to those of the original study are achievable.



The data annotation process was carried out by six volunteer annotators. Annotators were required to assign a label to each sentence. The labels were: Positive, Negative, Neutral, Mixed feelings and Non-Malayalam.

An example item from the corpus is shown here:

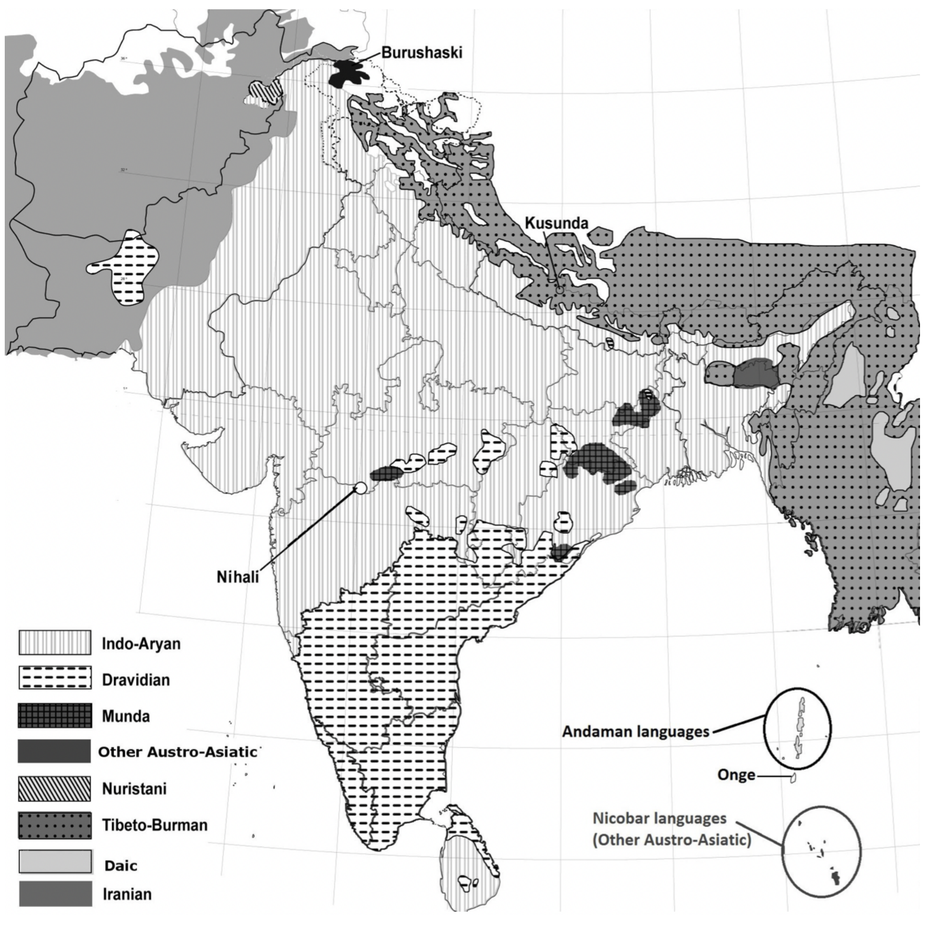
Ufff vere level ikkaaa ingha pwoli aahn
Another level, ikka you are awesome
Assigned Label: ***Positive***

- **Evaluate** the reproducibility of original study based on the **performance** of Machine Learning Models utilising different metrics.

Research Questions

- **RQ1**: How does the **Inter-Annotator Agreement** in the Re-Annotated Corpus Compare to the Original as measured by **Krippendorff's Alpha**?
- **RQ2**: Does the Introduction of New Labels Impact the **Performance** of the **Classifier** as Measured by **Precision, Recall, F1-score, Support and CV***?

Code-Mixed Language



The Distribution of Languages in India

- **Malayalam** is a **Dravidian language**, primarily spoken in the southern part of India, which is distinguished by its **complex** and **rich phonetic and grammatical structures**.

Research Reproducibility

- **Comparative** corpus analysis of **Original** and **Re-annotated** corpus was conducted.
- Several significant **challenges** that underscore the complexities of **research reproduciblity** were identified.
- The **challenges** that were encountered were mainly:
  - **Data Preparation Issues** such as preprocessing inconsistencies, unclear annotation process, missing information
  - **Data Classification Issues** such as lack of clear guidelines and code
- **Reproducibility** assessed using **model performance** on new labels and **Quantitative metrics** such as **Krippendorff's Alpha** and **CV***
- **Identify** the best practices to ensure **reproducibility of research**

RQ1 Inter-Annotator Agreement

- Calculate **IAA** on the **re-annotated corpus** of the **reproduction study** utilising **Krippendorff's Alpha**
- Conduct an **Inter-Study Agreement** utilising corpus **labels** of the **original study** and the **reproduction study**
- Compare the **IAA** of the **original** and **re-annotated corpus**
- **Result 1**: **Lower** Krippendorff's Alpha for the IAA indicates **challenges in reproducing consistent annotations**.
- **Result 2**: **Lower** alpha for the Inter Study Agreement suggests **inconsistent labels within the labels of the two studies**.

RQ2 Classifier Performance

- **Reproduce** selected **machine learning models** of the **original study**
- **Assess** the **reproduced models** performance utilising metrics such as **precision, recall, f1-score, support and CV***.
- **Result 1**: A **decline** in **performance** with the **re-annotated corpus** was observed.
- **Result 2**: **Higher CV* values** of the reproduced **LR and BERT models** indicates **bad reproducibility**.

References

[1] A. Arora et al. "Computational Historical Linguistics and Language Diversity in South Asia". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by S. Muresan, P. Nakov, and A. Villavicencio. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 1396–1409. DOI: 10.18653/v1/2022.acl-long.99. URL: https://aclanthology.org/2022.acl-long.99.

[2] B. R. Chakravarthi et al. "A Sentiment Analysis Dataset for Code-Mixed Malayalam-English". English. In: *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*. Ed. by D. Beermann et al. Marseille, France: European Language Resources association, May 2020, pp. 177–184. ISBN: 979-10-95546-35-1. URL: https://aclanthology.org/2020.sltu-1.25.

Contact

ss2246 *at* hw.ac.uk
Sachin Sasidharan Nair - MSc Data Science student at Heriot-Watt University
T.Dinkar *at* hw.ac.uk
Tanvi Dinkar was supported by the EPSRC projects 'AISEC: AI Secure and Explainable by Construction' (EP/T026952/1) and 'Gender Bias in Conversational AI' (EP/T023767/1)
G.Abercrombie *at* hw.ac.uk
Gavin Abercrombie was supported by the EPSRC project 'Equally Safe Online' (EP/W025493/1)