CS 772: Regional Multi-lingual Negative Sentiment Detection in Online Media

Neelkamal Bhuyan – 180020058

Amitrajit Bhattacharjee – 180020014

Aadesh Anil Madnaik - 180020014

Team 10

Problem Statement

Demo - Sentiment value for different sentences

- Input: text in the form of a sentence or whole passage
- Output: Two subtasks
  - Classified into: Overtly, Covertly or Not Aggressive (OAG/CAG/NAG)
  - Also, gendered classification (GEN/NGEN) has been explored

Motivation

We have observed media and all forms of communication move to a digital format, over the internet. While this has had a positive impact by bringing the world together, it has several negative implications in the form of hate speech, aggression and misogyny spread. We hope to be able to identify the following in an automated manner.

Basic Paper(s)

Papers which form basis of our work

[1] Towards Sub-Word Level Compositions for Sentiment Analysis of Hindi-English Code Mixed Text

Aditya Joshi, Ameya Prabhu, Manish Shrivastava, Vasudeva Varma

https://www.aclweb.org/anthology/C16-1234/

[2] Aggression Identification in English, Hindi and Bangla Text using BERT, RoBERTa and SVM

Arup Baruah, Kaushik Das, Ferdous Barbhuiya, Kuntal Dey

https://www.aclweb.org/anthology/2020.trac-1.12/

Data

- **Description** - Developing a Multilingual Annotated Corpus of <u>Misogyny</u> and <u>Aggression</u>
  - From the book Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying (May 2020), European Language Resources Association (ELRA)
- **URL** – https://github.com/cozek/trac2020_submission/tree/master/data/hin

Data

Citation details:

```
@InProceedings{trac2-dataset,

author = {Bhattacharya, Shiladitya and Singh, Siddharth and Kumar, Ritesh and Bansal, Akanksha
and Bhagat, Akash and Dawer, Yogesh and Lahiri, Bornini and Ojha, Atul Kr.},

title = {Developing a Multilingual Annotated Corpus of Misogyny and Aggression},

booktitle = {Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying},

month = {May},

year = {2020},

address = {Marseille, France},

publisher = {European Language Resources Association (ELRA)},

pages = {158--168},

url = {https://www.aclweb.org/anthology/2020.trac2-1.25}

}
```

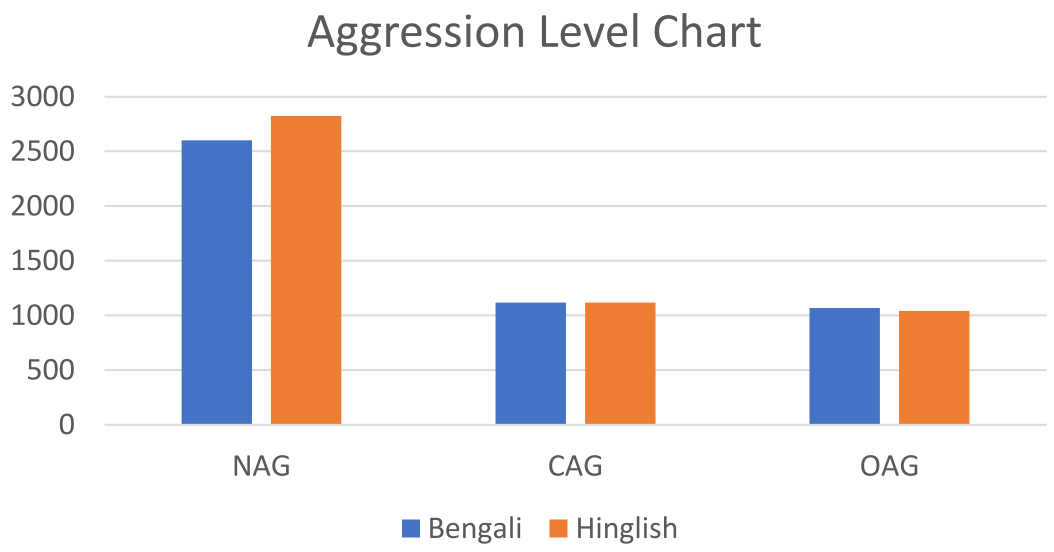
Data Statistics

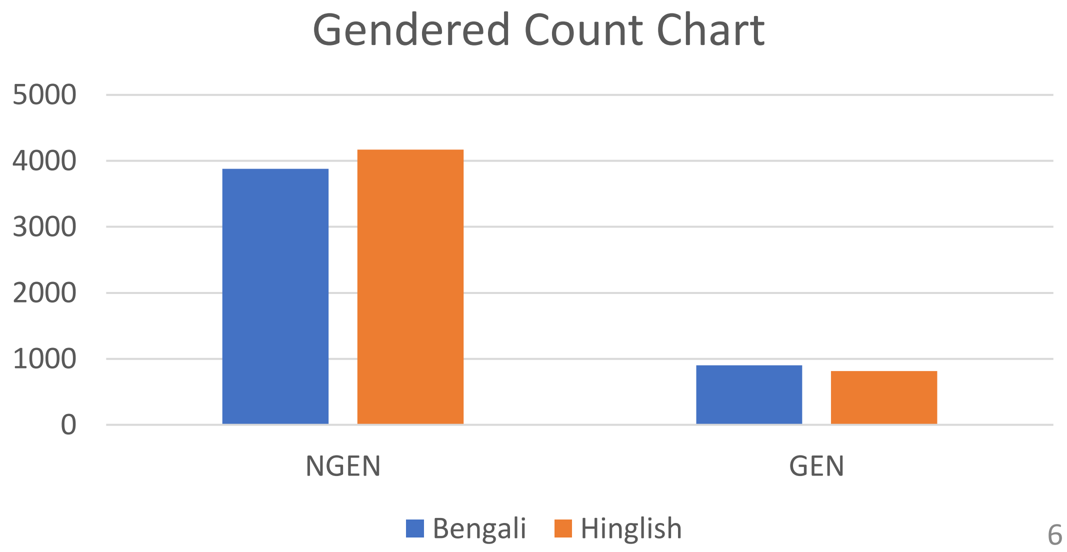
Bengali Dataset

```
Total dev + train size = 4783

NAG     2600
CAG     1116
OAG     1067
Name: Sub-task A, dtype: int64

NGEN     3880
GEN       903
Name: Sub-task B, dtype: int64
```

Hinglish Dataset

```
Total dev + train size = 4981

NAG     2823
OAG     1118
CAG     1040
Name: Sub-task A, dtype: int64

NGEN     4168
GEN       813
Name: Sub-task B, dtype: int64
```





Techniques used

Deep Learning & Transformers are the primary techniques

- XLM-RoBERTa transformer
  - Implementation using XLM-RoBERTa - outperforms M-BERT in multiple cross-lingual benchmarks
  - Batch size is set to 32
  - 270M parameters with 12-layers, 768-hidden-state, 3072 feed-forward hidden-state, 8-heads,
  - Trained on 2.5 TB of newly created clean CommonCrawl data in 100 languages

Advantages of RoBERTa over BERT

Few modifications to the original model such as (BERT)

- Training on a larger dataset
- Dynamically masking out tokens compared to the original static masking
- Used pre-trained base versions made available by the HuggingFace Transformers library

Description of XLM-RoBERTa

- XLM-RoBERTa (Conneau et al., 2019) is a cross-lingual model that aims to tackle 'the curse of multilinguality problem' of cross-lingual models
- trained on up-to 100 languages and out-performs M-BERT in multiple cross-lingual benchmarks
- Base version coupled with an attention head classifier
- This model is used in the sub-tasks of the Hindi and Bangla language

Transformer: XLM-RoBERTa

Bi-LSTM layer – 50 units – 0.1 dropout rate

GlobalMaxPool – 1D layer

Dense Layer – ReLU – 50 units

Dropout layer – Rate = 0.2

Dense layer – SoftMax – 2 or 3 units depending on subtask A or B

Results

Sub-task	Train accuracy	Validation accuracy	F-score	Precision	Recall
A	0.7358	0.6991	0.7305	0.6410	0.6234
B	0.8398	0.8475	0.8983	0.9142	0.8823

Confusion matrix for sub-task A

```
[[479  23  76]
 [ 40 120  48]
 [ 80  33  98]]
```

Confusion matrix for sub-task B

```
[[773  72]
 [103  49]]
```

Demo and Case Study

Qualitative / case studies - Hindi

1. Gendered sentences:

Example sentence	GEN probability	NGEN probability
Ladki pagal ho gayi hai	0.7785	0.2215
superb hai bhai. bohot accha kiya	0.1871	0.8129

2. Aggression-level analysis:

Example sentence	Not aggressive probability	Covertly aggressive probability	Overtly aggressive probability
Wo bohot hi sweet hai	0.8188	0.1230	0.0582
Do Chaar Murder hone k baad Hi Patta chalta hi ki Kaun Sachcha aur Jhootha	0.2389	0.4716	0.2895
Tere mar ja nay pai kisi ko fark nahi padta	0.1710	0.3980	0. 4309

Demo and Case Study

Qualitative / case studies - Bengali

1. Aggression-level analysis:

Example sentence	Not aggressive probability	Covertly aggressive probability	Overtly aggressive probability
Tumi thik bolcho	0.4515	0.2323	0.3160
Tar jonno payer juto mathay uthano jay na	0.2824	0.3704	0.3471
Aaami tomaake ekdum jore maarbo	0.2947	0.3465	0.3587

2. Gendered sentences:

Example sentence	GEN probability	NGEN probability
Jar jeta paowa uchit noy, seta take dite nei	0.3813	0.6187
Tumi paagol mohila	0.6485	0.3514

Conclusion and Future Work

In our work we have observed that XLM-RoBERTa performs well under Hinglish and Bengali texts. We identify two extensions for this model and training:

1. Speech-to-text for hate speech in videos on social media platforms
2. Extension through more multilingual transformers like

Results (Bengali)

Sub-task	Train accuracy	Validation accuracy	F-score	Precision	Recall
A	0.5572	0.5549	0.2709	0.3251	0.3497
B	0.8259	0.8004	0.8679	0.9217	0.8223

Confusion matrix for sub-task A

```
[[519    0    3]
 [208    0    9]
 [206    0   12]]
```

Confusion matrix for sub-task B

```
[[706  60]
 [155  36]]
```