LREC 2022 Marseille

MemoSen: A Multimodal Dataset for Sentiment Analysis of Memes

Eftekhar Hossain, Omar Sharif, and Mohammed Moshiul Hoque
Chittagong University of Engineering and Technology, Bangladesh

Introduction

➢ Sentiment analysis of memes has become a crucial research issue in low resource languages like Bengali.

Necessity

➢ To mitigate the spread of negativity and understand the public expression towards an event or topic.

Scarcity of benchmark corpora in Bengali

Challenges & Contribution

Challenging for the machines and humans for several reasons

- Memes are context dependent
- Visual and textual information are often disparate
- Embedded text is too short

Extracting the code-mixed and code switched text from the memes

✓ Created the ***MemoSen***, a multimodal sentiment analysis dataset for Bengali
✓ Annotated with Positive, Negative, Neutral labels.

Performed extensive experiments with state-of the-art visual and textual and multimodal models.

Class	Train	Test	Valid	Total
Positive	950	285	114	1349
Negative	2001	524	203	2728
Neutral	195	64	32	291

Table 1. Train, test, validation split

	Positive	Negative	Neutral
Positive	-	**0.355**	0.213
Negative	-	-	0.228

Table 2. Jaccard similarity of 400 most frequent words

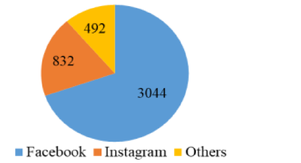
MemoSen: a New Benchmark Dataset

➢ ***MemoSen*** consists of 4368 memes.
➢ Considered memes with captions in Bengali, Bengali and English (code-mixed) or in Banglish (code-switched) manner.

✓Captions are manually extracted.

Manual labelling into-
✓Positive
✓Negative
✓Neutral

A mean kappa score of 0.674 is obtained between the three annotators



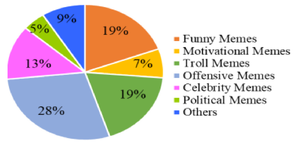


Figure 1. Source statistics of the MemoSen dataset

(a) meme shows affection (b) meme shows accolade (c) meme shows funny humor
(d) insult a person (e) denigrate a group of celebrities (f) shows obscene content

(g) memes with inherent sentiment (h) memes intention is incomprehensible

Figure 2. Few example memes from MemoSen: here (a,b,c) are the positive memes, (d,e,f) are the negative and (g,h) are the neutral memes

Methodology

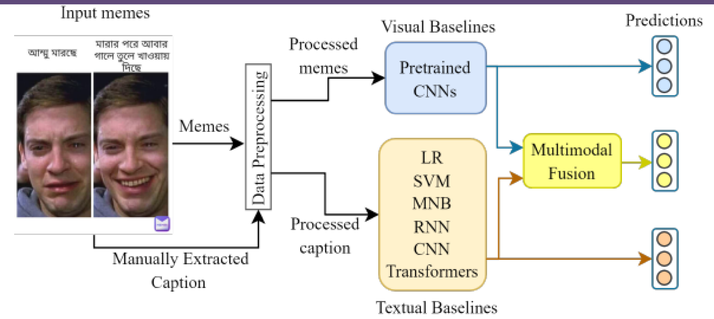


Figure 3. Abstract view of the Bengali meme sentiment classification system

MemoSen: Benchmark Evaluation

Approach	Models	P	R	WF
Visual	Xception	0.587	0.615	0.579
	VGG19	0.588	0.543	0.563
	VGG16	0.582	0.571	0.559
	ResNet50	0.602	0.628	0.600
	DenseNet	0.585	0.609	0.594
Textual	LR	0.617	0.663	0.608
	MNB	0.643	0.663	0.628
	SVM	0.670	0.653	0.608
	BiLSTM (B)	0.587	0.604	0.594
	CNN (C)	0.605	0.600	0.594
	B+C	0.606	0.554	0.576
	MuriL	0.624	0.640	**0.631**
	Bangla-BERT	0.622	0.605	0.605
	XLM-R	0.360	0.600	0.450

Table 3. Performance comparison of visual and textual models

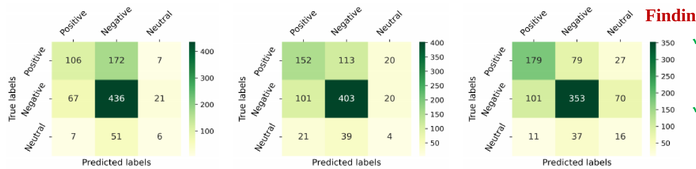
✓Textual model (MuriL) obtained the highest WF of **0.631**

		Models	P	R	WF
FF	R+	BiLSTM	0.625	0.633	0.626
		CNN	0.575	0.591	0.582
		BiLSTM+CNN	0.615	0.578	0.592
		MuriL	0.525	0.392	0.419
		Bangla-BERT	0.510	0.557	0.508
DF	R+	BiLSTM	0.644	0.631	0.635
		CNN	0.663	0.628	**0.643**
		BiLSTM+CNN	0.566	0.592	0.575
		MuriL	0.552	0.554	0.543
		Bangla-BERT	0.504	0.394	0.329

Table 4. Performance comparison of multimodal models on test

✓ResNet50+CNN outperformed the unimodal models with WF of **0.643**

Error Analysis



(a) Best Visual (ResNet50) (b) Best Textual (MuriL) (c) Best Multimodal (ResNet50+CNN)

Findings,

✓**visual information** is more appropriate for predicting the negative sentiment
✓**visual and textual** information is required to get a robust inference for the positive and neutral sentiment

Possible reasons of misclassification,

✓large number of words are **overlapped** between the classes
✓the presence of code-mixed and code-switched words
✓consistent visual features

Conclusion

➢ Developed ***'MemoSen'*** multimodal benchmark dataset
➢ Performed baseline evaluation with unimodal and multimodal features
➢ Multimodal approach achieved highest **weighted f1-score 0.643**