Paper Title: SentiGOLD: A Large Bangla Gold Standard Multi-Domain Sentiment Analysis Dataset and its Evaluation

Paper Link: https://arxiv.org/ftp/arxiv/papers/2306/2306.06147.pdf

1 Summary:

1.1 Motivation:

The primary motivation behind the text is to address the existing gap in Bangla sentiment analysis datasets. Despite the significance of sentiment analysis in various applications such as social media insight, election campaigns, and market research, there is a lack of large-scale, diverse, and reliable datasets for sentiment analysis in the Bangla language. The authors highlight the limitations of existing datasets, including issues with size, diversity, and annotation quality. The motivation of the paper lies in highlighting the lack of such resources in Bangla, especially due to the absence of a universally agreed-upon linguistic framework. The paper aims to fill this gap by introducing SentiGOLD, a large Bangla multi-domain sentiment analysis dataset comprising 70,000 samples from various sources. The motivation is further fueled by the need for a reliable dataset for sentiment analysis in Bangla, which is crucial for developing accurate and effective sentiment analysis models.

1.2 Contribution:

The paper presents the SentiGOLD dataset, is notably larger (approximately 4.5 times) than previous Bangla sentiment analysis resources and covers 30 domains, providing a substantial improvement in linguistic resources for sentiment analysis in Bangla. The development of SentiGOLD involves the implementation of a novel annotation procedure, enhancing the annotation quality. The annotation process is conducted under the approval of a national linguistics committee, ensuring a high Inter Annotator Agreement. The investigation involves different architectures and training methodologies, demonstrating the generalization capability of the dataset and achieving a 0.61 macro F1 score for 3 classes.

1.3 Methodology:

Initial Approach was to use feature models: unigram, bigram, trigram, and different combinations. Then, moved to different combinations of BiLSTM, Hierarchical Attention Network(HAN), biLSTM CNN with Attention(BCA) models. Lastly, moved to BERT-based pretrained language models. We adopted pretrained language models such as mBert and BanglaBert.

1.4 Conclusion:

The paper discusses the introduction of the SentiGOLD dataset to enhance linguistic resources for Bangla sentiment analysis. The development process incorporates a novel annotation procedure aimed at improving annotation quality, with potential applicability to other South Asian languages. The authors express their commitment to ongoing work, exploring different modeling architectures and incorporating diverse data from additional domains to enhance solutions. The contribution of SentiGOLD is positioned as a competitive benchmark, encouraging further research and improvements within the community.

2 Limitations:

2.1 First Limitation

Limited Generalization to Informal Texts: The fine-tuned model, BanglaBert, faces challenges in making correct predictions when exposed to data samples containing humorous/sarcastic words, politically controversial terms, provincial vitriols, and indirect attacking words, indicating a potential lack of generalization to noisy or informal language.

2.2 Second Limitation

Incomplete Training Dataset:
The SentiGold training dataset used for fine-tuning may not cover all possible witty or harsh words, suggesting a limitation in the diversity of training data.

3 Synthesis:

The paper introduces the SentiGOLD dataset, a substantial contribution to Bangla sentiment analysis,is meticulously annotated by a gender-balanced team of linguists, adhering to linguistic conventions established through collaboration with national stakeholders. SentiGOLD addresses the absence of standard sentiment analysis datasets in Bangla, providing a comprehensive resource that maintains rigorous domain and class distribution. The paper emphasizes the importance of a standard dataset with key characteristics, such as size, diversity, and high IAA score, for reliable sentiment analysis. The authors conduct thorough evaluations, including intra- and cross-dataset testing, using the SentNoB dataset. The SentiGOLD dataset's generalizability is demonstrated through zero-shot experiments, showcasing competitive performance with a top model achieving a macro F1 of 0.62 for 5 classes intra-dataset and 0.61 for 3 classes in cross-dataset testing, comparable to the current state-of-the-art. These contributions include proposing the dataset, ensuring high IAA, building an annotation management system, investigating different architectures, and achieving a 6% performance improvement over SentNoB. The authors openly share their dataset and annotation management system, fostering transparency and encouraging further research in Bangla sentiment analysis. The paper concludes by emphasizing the need for large-scale, diverse Bangla sentiment analysis datasets and positions. SentiGOLD as a benchmark, paving the way for advancements in the field. In future, these can be very helpful while working on Bangla sentiment analysis. The creation of SentiGOLD is driven by the desire to establish a benchmark for sentiment analysis in Bangla, providing researchers and practitioners with a standardized dataset for developing and evaluating sentiment analysis models across different domains.