Improving sentiment classification in low-resource Bengali language utilizing cross-lingual self-supervised learning

Salim Sazzed

Old Dominion University, Norfolk VA 23508, USA
ssazz001@odu.edu

Abstract. One of the barriers of sentiment analysis research in low-resource languages such as Bengali is the lack of annotated data. Manual annotation requires resources, which are scarcely available in low-resource languages. We present a cross-lingual hybrid methodology that utilizes machine translation and prior sentiment information to generate accurate pseudo-labels. By leveraging the pseudo-labels, a supervised ML classifier is trained for sentiment classification. We contrast the performance of the proposed self-supervised methodology with the Bengali and English sentiment classification methods (i.e., methods which do not require labeled data). We observe that the self-supervised hybrid methodology improves the macro F1 scores by 15%-25%. The results infer that the proposed framework can improve the performance of sentiment classification in low-resource languages that lack labeled data.

Keywords: Bangla sentiment analysis · pseudo-label generation · Cross-lingual sentiment analysis.

1 Introduction

Sentiment analysis determines the semantic orientation of an opinion expressed in a text. The rapid growth of user-generated online content necessitates analyzing user's opinions and emotions in textual data for various purposes. Researchers applied both the machine learning-based [17, 1] and lexicon-based methods [26] to classify sentiments at various levels of granularity such as binary, 3-class, or 5-class. The supervised ML methods usually exhibit much better performance; however, they require a large volume of annotated data.

English and several other languages enjoy ample resources, such as annotated data for sentiment analysis; however, such resources are not available in resource-constrained languages. The self-supervised approaches can be an effective way to deal with the inadequacy of labeled data in low-resource languages. Instead of manual annotation, the self-supervised learning methods automatically generate pseudo-labels by implicitly learning underlying patterns from the data or utilizing a set of rules.

Cross-lingual sentiment classification is another way to deal with resource scarcity issues in low-resource languages. Cross-lingual sentiment classification aims to leverage resources like labeled data and opinion lexicons from a resource-rich language (typically English) to classify the sentiment polarity of texts in a low-resource language. Though cross-lingual approaches have been studied in several low-resource languages [16, 5], in Bengali only a few works utilized it for sentiment classification [22] or sentiment lexicon creation [8]. In [22], the performances of various supervised ML classifiers have been compared in a Bengali corpus and corresponding machine-translated English version. The authors found Bengali-English machine translation system had reached some level of maturity; thus could be utilized for cross-lingual sentiment analysis.

In this work, we present a cross-lingual self-supervised methodology for classifying sentiments in unlabeled Bengali text. The proposed self-supervised hybrid methodology combines lexicon-based and supervised ML-based methods. Employing machine translation, we first transform Bengali text to English. Then we leverage prior word-level sentiment information (i.e., sentiment lexicon), a set of rules, and consensus-based filtering to generate accurate pseudo-labels for training a supervised ML classifier. We compare the performance of the proposed method with English lexicon-based sentiment analysis tools, VADER [13], TextBlob[1], and SentiStrength [24] and a Bengali lexicon-based method [21]. We observe that the hybrid approach improves the F1 score by 15% and accuracy by 11% compared to the best lexicon-based method.

1.1 Contributions

The major contributions of this work can be summarized as follows:

- We conduct a comparative performance analysis of Bengali and English lexicon-based methods.
- To elevate the performance of sentiment classification in unlabeled data, we present a cross-lingual self-supervised learning approach.
- We demonstrate how to generate highly accurate pseudo-labels to deal with the lack of labeled data in Bengali.
- We show that by utilizing machine translation and combining lexicon-based and ML-based methods, substantially improved performance can be attained.

2 Related Work

Most of the research in sentiment analysis has been conducted in English and a few other major languages such as Chinese, Arabic, and Spanish. In Bengali, limited research has been performed using corpora collected from various sources such as microblogs, Facebook statuses, and other social media sources [18, 9]. Researchers utilized various supervised methods, such as SVM with maximum entropy [7], Naive Bayes (NB) [14], Deep Neural Network [25, 11] for Bengali sentiment analysis. The word-embedding-based approach has been explored in [2].

Cross-lingual approaches of sentiment classification have been applied to several low-resource languages. The linked WordNets was used in [4] to bridge the

[1] https://textblob.readthedocs.io/en/dev/

language gap between two Indian languages, Hindi and Marathi. The performance and effectiveness of machine translation systems and supervised methods for multilingual sentiment analysis was investigated in [3] using four languages: English, German, Spanish, and French; three machine translation systems: Google, Bing, and Moses; several supervised algorithms and various types of features. [16] proposed a cross-lingual mixture model (CLMM) that exploits unlabeled bilingual parallel corpus. In [5], authors utilized a machine translation system for projecting resources from English to Romanian and Spanish and obtained a comparative performance. In [10], the authors proposed an end-to-end cross-lingual sentiment analysis (CLSA) model by leveraging unlabeled data in multiple languages and domains. The authors of [27] proposed a learning approach that does not require any cross-lingual labeled data. Their algorithm optimizes the transformation functions of monolingual word-embedding space. The authors of [6] introduced an Adversarial Deep Averaging Network (ADAN) that uses a shared feature extractor to learn hidden representations that are invariant across languages. Their experiments on Chinese and Arabic sentiment classification demonstrated the efficacy of ADAN.

The cross-lingual approach of sentiment analysis in Bengali is still largely unexplored; only a few works investigated it for tasks such as translating English polarity lexicon to Bengali [8], comparing the performance of ML algorithms in Bengali and machine-translated corpus [22]. The authors of [22] utilized two small datasets to compare the performance of supervised ML algorithms in Bengali and machine-translated English corpora. They found supervised ML algorithms showed better performance in the model trained on the translated corpus.

A plethora of studies explored hybrid approaches of sentiment classification; however, most of them utilized labeled or partially labeled datasets. A hybrid method was proposed in [28] for sentiment analysis in Twitter data that does not require any labeled data. The proposed method adopted a lexicon-based approach to label the training examples. The authors of [12] proposed a framework where an initial classifier is learned by incorporating a sentiment lexicon and using generalized expectation criteria. SESS (SElf-Supervised and Syntax-Based method) [29] works in three phases; initially, some documents are classified iteratively based on a sentiment dictionary. Afterward, a machine learning model is trained using the classified documents, and finally, the learned model is applied to the whole data set.

To the best of our knowledge, this work is the first attempt to incorporate the cross-lingual setting with self-supervised learning in Bengali. Compared to the existing self-supervised approaches, the proposed methodology differs in the way we perform pseudo-label generation and selection, training-testing set split, and model training.

3 Dataset and Machine Translation

3.1 Dataset

We use a large annotated review corpus[2] deposited by the author of [20]. The reviews in the corpus represent viewer's opinions toward a number of Bengali dramas. The data collection and annotation procedures were described in [20].

Bengali Reviews	Machine Translation	Polarity
সত্যিই দুর্দান্ত দুর্দান্ত উভয় অংশ ... আমি একজন ভারতীয় ... তবে আমি বাংলাদেশী নাটক পছন্দ করি। বিশেষভাবে জিয়াউল ফারুক অপূর্ব স্যার এর ... আমি আপনাকে স্যার ভালবাসি	Both really awesome great parts ... I'm an Indian ... But I like Bangladeshi drama. In particular Ziaul Farooq is a wonderful sir ... I love you sir.	Positive
এই নাটকটি আমার জীবনের সাথে জড়িয়ে আছে।সত্যি কিছু কিছু স্মৃতি থেকে যায়, যা কখনো ভুলা যায় না।এই নাটকটি আমার ভীষণ ভালো লেগেছে! সবাই ভালো থাকবেন।	This drama is embedded in our lives.There are some memories that will never be forgotten.This drama is very like ours! Everyone will be fine.	Positive
আমি কোনো এতো হাসি নাই,যতটা না এই নাটক দেখে হাসছি।পুরো টা সময় শামীম ভাইয়ার অভিনয় উপভোগ করলাম	I didn't laugh at all, until I was laughing at this drama.I enjoyed Shamim Bhai's acting the whole time	Positive
বাংলাদেশের নাটক কেমন যেন দিন দিন দেখার অযোগ্য হয়ে পড়ছে। এসব নাটক পরিবারকে নিয়ে দেখার মত নয়।কি সব ধরনের নাম আর কি সব ভাষা ব্যবহার করে নিজেই বুঝিনা।এইসব নাটক থেকে আমরা কি শিক্ষা নিতে পারি। আমাদের সমাজ পরিবার দেশ সব কিছু ধ্বংস করে দিচ্ছে সব ধরনের কিছু নাটক। এসব পরিচালনা দের জুতাপেটা করে বাংলাদেশ থেকে বিতাড়িত করা হোক।	How is the drama of Bangladesh becoming inaccessible day by day. These plays are not worth watching with the family.Do not use all kinds of names and all the languages myself.What can we learn from these plays? Our society family country is destroying everything drama of all kinds. Be expelled from Bangladesh by wearing these shoes.	Negative
নাটকের কোন কনসেপ্ট নেই, কোনো কাহিনী নেই... অদ্ভুত উটের পিঠে চলছে স্বদেশ।	No Concept of Drama, No Story ... Strange camels running the country.	Negative
বাল আমার।।। ফালতু স্টোরি, ফালতু মেকিং। এসব আজাইরা নাটকের জন্যই দিন দিন বাংলা নাটকের মান কমে যাচ্ছে।	My boy. False Story, False Making. The quality of Bangla drama is declining day by day for these Azira plays.	Negative

Fig. 1. Example of Bengali and translated English reviews with annotations

This review corpus consists of 11807 annotated reviews, where each review contains between 2 to 300 Bengali words. This class-imbalanced dataset comprised of 3307 *negative* and 8500 *positive* reviews. From the annotator ratings, the author observed an inter-rater agreement of around 0.83 based on Cohen's κ. The reviews are highly polar since reviews that are marked as non-subjective by either of the annotators were excluded. Figure 1 shows some examples of Bengali reviews and corresponding English machine translation with annotations.

3.2 Quality of Machine Translation and Sentiment Preservation

To leverage cross-lingual resources, it is required to link the source and target languages. The machine translation (MT) service is one of the most prevalent ways to connect languages. The quality of a machine translation system largely depends on the amount of training data used for model training. Without using an advanced machine translation service built on a huge training dataset, good translation accuracy is not attainable. The authors of [22] utilized Google Translate [3] to translate Bengali reviews to English for cross-lingual sentiment analysis. They manually assessed the quality of the machine translation and observed that the quality of the translation varies among reviews. Among 1016 translated Bengali comments, on a Likert scale of 1-5, they assigned 170 comments to a rating of 1, 279 comments as 2, 229 comments as 3, 140 comments as

[2] https://github.com/sazzadcsedu/BN-Dataset.git
[3] https://translate.google.com

4, and 198 comments as 5, with an average translation rating of 2.92, which they described as fair. Therefore, in this work, we use Google Translate to translate the Bengali reviews into English.

To investigate the sentiment preservation after machine translation, the author of [20] computed the agreement of the predictions of two highly accurate ML classifiers, logistic regression (LR) and support vector machine (SVM) in Bengali and machine-translated English corpus in a drama review dataset. The author utilized Cohen's kappa and Gwet's AC1 to assess inter-rater agreements. Both SVM and LR show kappa scores above 0.80 and AC1 scores above 0.85 (where a score of 1 refers to perfect agreement). The results indicate sentiment consistency exists between original Bengali and machine-translated English reviews.

The above-mentioned studies suggest that the quality of Bengali-to-English machine translation is fair, and the sentiment is preserved in most cases. Therefore, no manual error correction is employed in the machine-translated reviews. Besides, one of the main objectives of this work is to eliminate manual intervention.

4 Cross-lingual Sentiment Analysis in Bengali

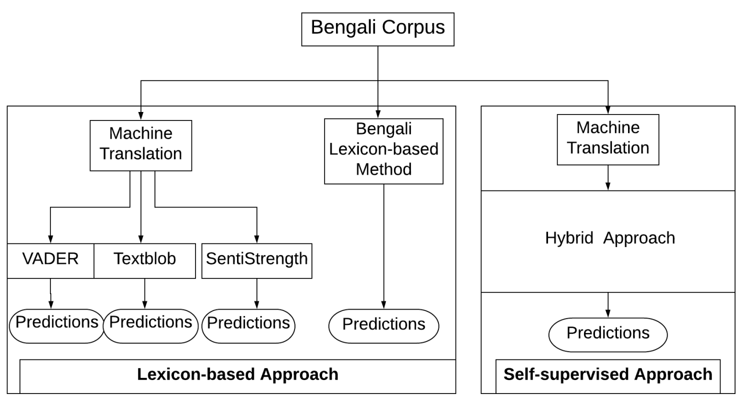

Fig. 2. Various approaches of sentiment analysis in Bengali review corpus

4.1 Lexicon-based Methods

To find the efficacy of the lexicon-based methods for sentiment classification in the translated corpus, three popular lexicon-based tools, SentiStrength, VADER, and TextBlob are employed. A non-negative polarity score of SentiStrength and TextBlob refers to a positive class prediction; otherwise, we consider it as a negative prediction. For VADER, the compound score is used instead of the polarity score.

For Bengali, we utilize a publicly available Bengali sentiment lexicon [21] and a set of linguistic rules. This binary-weighted lexicon consists of around 700 opinion words, where positive and negative words have a weight of +1 and -1, respectively. Besides applying the word-level polarity, we employ a simple negation rule to address the shift of polarity. The class assignment based on the review polarity score is implemented similarly to English lexicon-based methods. In Bengali, only a few works employed the lexicon-based methods for sentiment classification due to a lack of standard language-specific resources (e.g., sentiment lexicon, POS tagger, dependency parser, etc.). Besides, their implementations are not publicly available.

4.2 Self-supervised Hybrid Methodology

Self-supervised learning utilizes data that is automatically labeled by learning patterns, exploiting the relationships between features, and employing rules. As the Bengali lexicon-based method [21] yields comparatively lower accuracy, we integrate an English lexicon-based method [23] in the self-supervised framework for automatically generating labels. The steps of the proposed methodology are shown in Fig. 3.

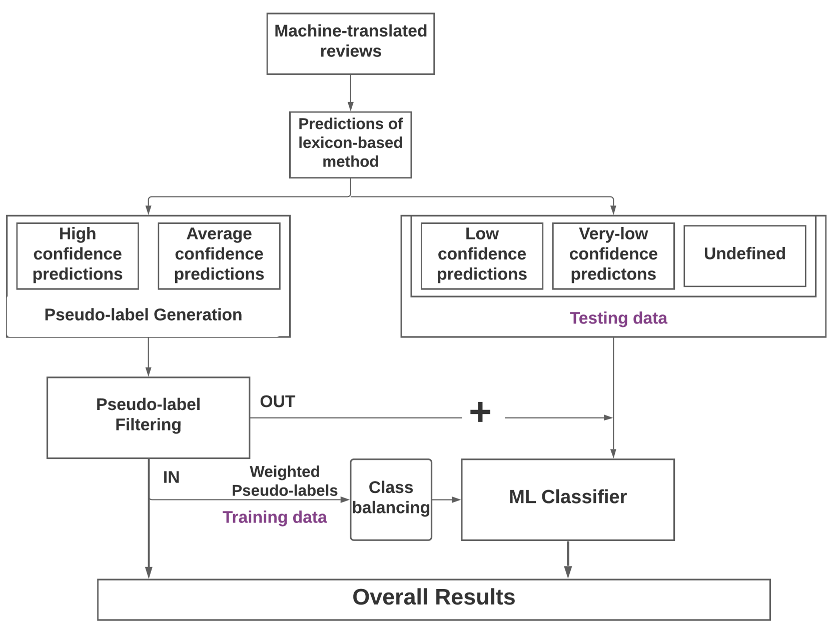

Fig. 3. Steps of the proposed self-supervised hybrid methodology

The proposed cross-lingual self-supervised method trains an ML classifier by following several steps. First, the Bengali reviews are translated to English utilizing Google Translate. Then we employ an English lexicon-based method

[23] to generate highly accurate pseudo-labels, which are used as a training set for ML classifiers. Afterward, a filtering step is applied to remove some of the pseudo-labeled reviews from the training set. In the final step, weights are assigned to the filtered pseudo-labels to train a supervised ML classifier.

Pseudo-label Generation. To generate highly accurate pseudo-labels for the supervised ML classifier, we utilize a lexicon-based method, LRSentiA [23]. In addition to determining the semantic orientation of a review, LRSentiA provides the confidence score of the prediction. The prediction confidence score $ConfScore(r)$ of a review r is determined using the following equation-

$$ConfScore(r) = \frac{abs(P_{pos}(r) + P_{neg}(r))}{abs(P_{pos}(r)) + abs(P_{neg}(r))}$$

As the equation indicates, the confidence score of the review r, $ConfScore(r)$, depends on the positive terms, $P_{pos}(r)$ and negative terms, $P_{neg}(r)$ present in the review. A large presence of either positive or negative terms indicates a high confidence score. When the lexicon-based method predicts the class of a review with high confidence, then it is a highly polar review, as indicated by the above equation. These highly polarized reviews have a low chance for misclassification; thus can be used as pseudo-labeled training data.

Table 1. Prediction accuracy of LRSentiA across various confidence groups in machine translated corpus

Confidence Group	ConfScore	Accuracy	#Review
high	(0.75, 1.0]	98.1%	7596
average	(0.5, 0.75]	91.2%	1609
low	(0.25, 0.5]	84.3%	633
very-low	(0.0, 0.25]	79.5%	462

Based on the prediction confidence scores of n reviews, $ConfScore(r_1)$,...., $ConfScore(r_n)$, we categorize them into five confidence groups (n equals to the number of reviews in the corpus). The reviews with a confidence score above 0.75 belong to *high* confidence category, reviews having confidence score between (>0.5) and 0.75 belong to *average* confidence category, between (>0.25) and 0.5 fall into *low* category, between (>0) and 0.25 fall into *very-low* category and remaining reviews with 0 confidence score fall into *undefined* group.

Three criteria are considered, similar to [23], while categorizing predictions into multiple confidence groups that are described below.

[a]. Minimizing the inclusion of wrong predictions (i.e., inaccurate pseudo-labels) in the training set to restrict error propagation to the classifier.
[b]. Maximizing the number of reviews (i.e., a large training set) included in the training data.

[c]. Show the correlation between the prediction confidence score and the accuracy (i.e., prediction with a high confidence score implies correctness) to satisfy both criteria [a] and [b].

Both [a] and [b] assist in achieving better performance from the ML classifier. The highly accurate pseudo-labels ([a]) imply less error-propagation to the classifier, and a higher number of pseudo-labels ([b]) mean a large training set, which is required to get good performance from the ML model. [c] helps to determine which reviews should go to training data and which ones to be used as testing data.

We observe that discretizing the reviews into five categories satisfies all the criteria (i.e., [a], [b], and [c]) best; therefore, five confidence groups are used. Table 1 shows accuracies of different confidence groups. The results suggest that there exists a correlation between the prediction accuracy and confidence scores.

Pseudo-label Filtering. This step involves filtering out some of the pseudo-labels selected from *high* and *average* confidence groups. The goal is to improve the accuracy of pseudo-labels further that are used in the training process. We apply the consensus-based filtering based on the lexicon-based method and SVM classifier. We perform 10-fold cross-validation utilizing these pseudo-labeled data from *high* and *average* confidence groups. Based on the predictions of SVM, we only keep the reviews that are assigned to the same class by both SVM and the lexicon-based method. These reviews are utilized as training data for the proposed self-supervised method. The discarded reviews are added to the testing data along with the reviews from *low*, *very-low* and *undefined* categories. The default parameter settings of scikit-learn library [19] and unigram and bigram based tf-idf features are used for the SVM classifier.

After the filtering step, we find 7082 reviews belong to *high confidence* group with an accuracy of around 98.5%. Since the accuracy of this group is already high, the improvement is not significant. However, for the next confidence group,*average*, we observe improvement in the accuracy from 91.2% (1609 reviews) to 93.7% (1321 reviews).

Pseudo-label Weighting. In this step, we assign the weights of the pseudo-labels. We calculate the average confidence scores of *high* and *average* confidence groups. Based on the average confidence scores, we assign the weights of the pseudo-labels that are used as training data for the ML classifiers. The reviews belong to *high* confidence group have higher weight compared to *average* confidence group. The weights of the pseudo-labels (i.e., influence to the classifier) are set based on the group confidence score instead of its own confidence score, as it is a more flexible measure.

Model Training. As shown in Table 1, the *high* and *average* confidence categories of the lexicon-based method yield mostly accurate predictions and can be used as pseudo-labels for the supervised ML algorithms. However, we observe that the distributions of *high* and *average* confidence groups are biased toward *positive* class, contain a much higher number of *positive* samples (could also be attributed to the class distribution of the original dataset). The performance of a supervised ML algorithm can be affected by the presence of the class imbalance. To reduce the negative impact of class inequality, we apply a sampling algorithm, Synthetic Minority Over-sampling TEchnique (SMOTE) [15]. SMOTE is an oversampling method that creates synthetic minority class samples. However, this sampling technique was not able to eliminate the bias towards the *positive* class in our experiment.

Therefore, we use a balanced subset from the *high* and *average* prediction categories to train the supervised ML classifier. The number of instances of each class in the subset is determined by the minimum value of the positive class instance and negative class instance. The instances of the dominant class are randomly selected. The results reported here are the average of the results of 10 random selection.

The reviews belong to *low*, *very-low*, and *undefined* prediction categories in which the lexicon-based method yield low accuracy and the discarded reviews in pseudo-label filtering step are used as testing data for the supervised ML classifiers. We extract unigram and bigram word features from the reviews, calculate the tf-idf scores and feed the scores to the machine learning classifiers. We use the default parameters settings of scikit-learn library [19] for all the ML classifiers.

Overall Predictions. As described above, the overall predictions of the hybrid methodology is determined by the combined predictions of the lexicon-based method (i.e., for reviews belong to *high* and *average* confidence groups excluding filtered out reviews) and the ML classifier (i.e., *low*, *very-low*, and *undefined* confidence groups plus filtered out reviews). The lexicon-based method successfully classifies reviews that are highly polarized (i.e., belong to *high*, and *average* confidence categories), with an accuracy of above 90%. However, for less polar and hard-to-distinguish reviews, the lexicon-based method shows lower accuracy due to various reasons (e.g., the polarity of a review is not obvious or lexicon-coverage problems). Therefore, for reviews belong to these groups, we utilize an ML classifier that is more robust for classifying complicated cases.

5 Results and Discussion

We compare the performances of various classifiers in the Bengali corpus and its machine-translated English version utilizing accuracy, precision, recall, and macro F1 score. The results of ML classifiers are reported based on the default parameter settings of the scikit-learn library [19].

Table 2 shows the performances of the lexicon-based methods in Bengali and translated English corpus. VADER and TextBlob exhibit similar F1 scores and accuracies, while SentiStrength performs relatively worse. VADER achieves an F1 score of 0.771 and an accuracy of 82.56%, while TextBlob obtains 0.776 and 82.79%, respectively. The Bengali lexicon-based method shows an F1 score of 0.699 and an accuracy of 77.10%.

Table 3 shows the performance of the self-supervised hybrid approach in the machine-translated corpus. The best F1 score of 0.897 is achieved when either LR (Logistic Regression) or SVM (Support Vector Classifier) classifier is integrated

Table 2. The performances of lexicon-based classifiers in Bengali and machine-translated English corpus.

Language of Corpus	Method	Precision	Recall	F1 Score	Accuracy
Translated English	VADER	0.846	0.707	0.771	82.56%
	TextBlob	0.863	0.705	0.776	82.79%
	SentiStrength	0.787	0.645	0.708	78.61%
Bengali	[21]	0.716	0.684	0.699	77.10%

Table 3. The performance of the proposed hybrid method in the machine-translated corpus integrating various ML classifiers

	Precision	Recall	F1 Score	Accuracy
Self-Supervised-Hybrid-SVM	0.891	0.903	0.897	91.5%
Self-Supervised-Hybrid-LR	0.876	0.919	0.897	90.8%
Self-Supervised-Hybrid-RF	0.888	0.858	0.872	90.0%
Self-Supervised-Hybrid-ET	0.888	0.872	0.880	90.5%

into the hybrid method. SVM provides the best accuracy of 91.5%. The decision tree-based methods RF (Random Forest) and ET (Extra Trees Classifier) achieve relatively lower F1 scores.

Among the three English lexicon-based methods applied to the translated reviews, TextBlob and VADER perform similarly, while SentiStrength shows relatively lower efficacy. Compared to English lexicon-based methods, the Bengali lexicon-based method exhibits inferior performance. Sentiment analysis research in Bengali is still not matured; therefore, it lacks enough resources. For example, in Bengali, no sophisticated and comprehensive sentiment lexicon exists. The sentiment lexicon we use here is small in size, consists of around 700 opinion words; thus, it lacks coverage of sentiment words, which is reflected in the performance of the Bengali sentiment analysis tool.

The self-supervised hybrid methodology improves the performance of the sentiment classification. Substantial improvements in both the F1 scores and accuracies are observed compared to the lexicon-based methods. As seen by Table 1, the confidence score of the prediction of the lexicon-based method is highly correlated with the prediction accuracy. The *high* category has a prediction accuracy of above 95% and *average* category has prediction accuracy of over 90%. Therefore, predictions from these categories can be used as training data for supervised ML classifiers with minimal negative impact. As low-resource languages suffer from data annotation issues, the proposed approach can boost sentiment analysis research in resource-poor languages.

6 Summary and Conclusions

In this work, we present a cross-lingual self-supervised methodology for improving the performance of sentiment classification in Bengali by automatically generating pseudo-labels. The proposed approach has advantages over the existing supervised classification methods, as it does not require manual labeling of reviews. As annotated data are hardly available in Bengali, and no sophisticated tools are available for sentiment analysis in unlabeled Bengali text, we explore the adaptation of resources and tools from English. We show that the hybrid cross-lingual approach substantially improves the performance of sentiment classification in Bengali. The results imply that the proposed methodology can advance sentiment analysis research in resource-constraints languages such as Bengali.

References

1. Abdi, A., Shamsuddin, S.M., Hasan, S., Piran, J.: Deep learning-based sentiment classification of evaluative text based on multi-feature fusion. Information Processing & Management **56**(4), 1245–1259 (2019)
2. Al-Amin, M., Islam, M.S., Uzzal, S.D.: Sentiment analysis of bengali comments with word2vec and sentiment information of words. In: 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE). pp. 186–190 (Feb 2017). https://doi.org/10.1109/ECACE.2017.7912903
3. Balahur, A., Turchi, M.: Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. Computer Speech Language **28**(1), 56 – 75 (2014). https://doi.org/https://doi.org/10.1016/j.csl.2013.03.004
4. Balamurali, A., Joshi, A., Bhattacharyya, P.: Cross-lingual sentiment analysis for indian languages using linked wordnets. In: COLING (2012)
5. Banea, C., Mihalcea, R., Wiebe, J., Hassan, S.: Multilingual subjectivity analysis using machine translation. In: 2008 Conference on Empirical Methods in Natural Language Processing. pp. 127–135 (2008)
6. Chen, X., Sun, Y., Athiwaratkun, B., Cardie, C., Weinberger, K.: Adversarial deep averaging networks for cross-lingual sentiment classification. Transactions of the Association for Computational Linguistics **6**, 557–570 (2018)
7. Chowdhury, S., Chowdhury, W.: Performing sentiment analysis in bangla microblog posts. In: 2014 International Conference on Informatics, Electronics Vision (ICIEV). pp. 1–6 (May 2014)
8. Das, A., Bandyopadhyay, S.: Sentiwordnet for bangla. Knowledge Sharing Event-4: Task **2**, 1–8 (2010)
9. Das, A., Bandyopadhyay, S.: Topic-based bengali opinion summarization. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters. pp. 232–240. Association for Computational Linguistics (2010)
10. Feng, Y., Wan, X.: Towards a unified end-to-end approach for fully unsupervised cross-lingual sentiment analysis. In: Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL). pp. 1035–1044. Hong Kong, China (Nov 2019)
11. Hassan, A., Amin, M.R., Al Azad, A.K., Mohammed, N.: Sentiment analysis on bangla and romanized bangla text using deep recurrent models. In: 2016 International Workshop on Computational Intelligence (IWCI). pp. 51–56. IEEE (2016)
12. He, Y., Zhou, D.: Self-training from labeled features for sentiment analysis. Information Processing & Management **47**(4), 606–616 (2011)
13. Hutto, C.J., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Eighth international AAAI conference on weblogs and social media (2014)

14. Islam, M.S., Islam, M.A., Hossain, M.A., Dey, J.J.: Supervised approach of sentimentality extraction from bengali facebook status. In: 2016 19th International Conference on Computer and Information Technology (ICCIT). pp. 383–387 (Dec 2016)
15. Lusa, L., et al.: Smote for high-dimensional class-imbalanced data. BMC bioinformatics **14**(1), 106 (2013)
16. Meng, X., Wei, F., Liu, X., Zhou, M., Xu, G., Wang, H.: Cross-lingual mixture model for sentiment classification. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1. pp. 572–581
17. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. pp. 79–86. Association for Computational Linguistics (2002)
18. Patra, B.G., Das, D., Das, A., Prasath, R.: Shared task on sentiment analysis in indian languages (sail) tweets-an overview. In: International Conference on Mining Intelligence and Knowledge Exploration. pp. 650–655. Springer (2015)
19. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
20. Sazzed, S.: Cross-lingual sentiment classification in low-resource bengali language. In: Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020). pp. 50–60 (2020)
21. Sazzed, S.: Development of sentiment lexicon in bengali utilizing corpus and cross-lingual resources. In: 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI). pp. 237–244. IEEE (2020)
22. Sazzed, S., Jayarathna, S.: A sentiment classification in bengali and machine translated english corpus. 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI) pp. 107–114 (2019)
23. Sazzed, S., Jayarathna, S.: Ssentia: A self-supervised sentiment analyzer for classification from unlabeled data. Machine Learning with Applications **4**, 100026 (2021)
24. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment strength detection in short informal text. Journal of the American society for information science and technology **61**(12), 2544–2558 (2010)
25. Tripto, N., Eunus Ali, M.: Detecting multilabel sentiment and emotions from bangla youtube comments. In: 2018 International Conference on Bangla Speech and Language Processing (ICBSLP). pp. 1–6 (09 2018)
26. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting on association for computational linguistics. pp. 417–424. Association for Computational Linguistics (2002)
27. Xu, R., Yang, Y., Otani, N., Wu, Y.: Unsupervised cross-lingual transfer of word embedding spaces. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. pp. 2465–2474. Brussels, Belgium (Oct-Nov 2018)
28. Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., Liu, B.: Combining lexicon-based and learning-based methods for twitter sentiment analysis. HP Laboratories, Technical Report HPL-2011 **89** (2011)
29. Zhang, W., Zhao, K., Qiu, L., Hu, C.: Sess: A self-supervised and syntax-based method for sentiment classification. In: Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation, Volume 2. pp. 596–605 (2009)