Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 2788–2795
Marseille, 11–16 May 2020

Annotated Corpus for Sentiment Analysis in Odia Language

Gaurav Mohanty, Pruthwik Mishra, Radhika Mamidi
Language Technologies Research Center
International Institute of Information Technology-Hyderabad
Telangana, India
{gaurav.mohanty,pruthwik.mishra}@research.iiit.ac.in, radhika.mamidi@iiit.ac.in

Abstract

Given the lack of an annotated corpus of non-traditional Odia literature which serves as the standard when it comes sentiment analysis, we have created an annotated corpus of Odia sentences and made it publicly available to promote research in the field. Secondly, in order to test the usability of currently available Odia sentiment lexicon, we experimented with various classifiers by training and testing on the sentiment annotated corpus while using identified affective words from the same as features. Annotation and classification are done at sentence level as the usage of sentiment lexicon is best suited to sentiment analysis at this level. The created corpus contains 2045 Odia sentences from news domain annotated with sentiment labels using a well-defined annotation scheme. An inter-annotator agreement score of 0.79 is reported for the corpus.

Keywords: Odia Corpus, Sentiment Analysis, Language Resource, Resource Creation

1. Introduction

Most of the research in sentiment analysis is focused on genres such as news data, customer reviews and tweets. Sentiment annotated corpus is useful to build models for the task of sentiment analysis. For these genres, annotation usually takes place at sentence and phrase level. Odia[1], being a resource-poor language, does not have such an annotated corpus available for public use. However, we have attempted to create an annotated corpus for Odia poetry, previously in literature (Mohanty et al., 2018).

In this paper, we create a sentiment annotated corpus of Odia sentences in News domain. News data may have opinionated references along with factual data. Hence at a sentence level these can be classified into positive, negative and neutral categories. Moreover, sentence level sentiment analysis provides room for usage of a sentiment lexicon for identifying affective words. We built an Odia sentiment lexicon for the task of sentiment classification previously (Mohanty et al., 2017). This lexicon has been built by using resources available for three other Indian languages: Bengali, Telugu, Tamil (Das and Bandyopadhyay, 2010) which are similar to Odia. IndoWordNet (Bhattacharyya, 2010) was used for establishing language pairs between Odia and each of the three aforementioned Indian languages. Classification performance using the Odia sentiment lexicon should provide valuable insight on the usability of this sentiment lexicon.

We have created an annotated corpus of Odia sentences from the abundantly available data in news domain for the language. This has further been made publicly available to promote research in the field. Secondly, in order to test the usability the already present Odia sentiment lexicon, we experimented with various classifiers by training and testing on the sentiment annotated corpus while using identified affective words from the same as features. Annotation and classification are done at sentence level as the usage of sentiment lexicon is best suited to sentiment analysis at this level. The created corpus contains 2045 Odia sentences from news domain annotated with sentiment labels. Furthermore, we have leveraged the vastly available data in news domain to compute Word Vector representations for Odia language. These can be used in the future as features for training models for the task of sentiment analysis.

[1] https://en.wikipedia.org/wiki/Odia_language

2. Data Collection

2.1. Source

Though reviews on e-commerce websites and customer feedback are best suited for the task of sentiment analysis, such data is not available in sufficiently large quantities over the Internet for Odia language. There is, however, enormous amount of data available in news domain in many Indian languages including Odia. News articles contain opinions mixed with neutral/factual statements. They are available in several news genres and serve as one of the standard corpus domains. Moreover, in the news domain in Odia, availability of data is vast.

For collecting news articles we used the *Samaaja* News Archive. The *Samaaja* News adds articles to its archive on a daily basis and hence serves as an excellent source of Odia data, not only for sentiment analysis, but also several other natural language processing related tasks. Hence, we use Odia news data from this source to be able to create word vector representations for the language. We briefly elaborate the procedure for extraction of these articles in the pre-processing step below.

2.2. Pre-processing

Before the actual data in the articles can be used for the task at hand, it needs to go through some amount

of pre-processing. It is to be noted that the data available in the Samaja News Archive Website [2] is available in an encoded format. A total of 100k articles were scrapped from the archive in the encoded format. These articles were then decoded to make the data available in Odia script (utf-8 encoding).
Each article contained several meta-data information which did not serve the task at hand. Meta-data included author of the article, date of publication, news location, and title of the article. Title of the article was kept as a part of the final dataset as this may contain sentiment information. The rest of the meta-data was removed. We used 175 articles from this dataset for the annotation task.
Sentence segmentation was carried out for every single article. Even though the focus was sentence level annotation, each sentence was numbered with it's corresponding article number and line number in order to help at document level analysis in the future. Classification of individual sentences in a given article may give a clearer picture of the overall sentiment of that article. This serves as a more granular option in comparison to direct document level sentiment classification.

3. Annotation

3.1. Annotation Scheme

Before annotation, a scheme was defined in order to help the annotators with labelling individual sentences. News domain contains several factual statements and sentences which do not have any positive/negative opinion to them as such. Hence, such sentences were classified as having neutral sentiment. Sentences were categorized into one of three classes: positive, negative or neutral.

Positive Sentences
These reflect positive opinion, emotion, feeling or sentiment such as expression of support, motivation, admiration, positive attitude, cheerfulness, forgiving nature, positive emotional state, etc. Positive sentences tend to have positive affective words present in them. A few examples are listed below.

- ସେ ଦେଶବିଦେଶରେ ଓଡ଼ିଶୀ ନୃତ୍ୟ ପରିବେଷଣ କରି ସଫଳତା ହାସଲ କରିଥିଲେ
  Transliteration: Se desabidesare Odissi nruthya paribesana kari saphalathaa haasala karithile
  Affective Words: ସଫଳତା, ହାସଲ
  English: She has achieved success by performing Odissi dance both nationally and internationally.

- ବୋଲିଂ ସହ ଭଲ ବ୍ୟାଟିଂ କରିବାରେ ପରେରା ସିଦ୍ଧହସ
  Transliteration: bowling saha bhala batting karibaare parera sidhahastha
  Affective Words: ଭଲ, ସିଦ୍ଧହସ୍ତ
  English: Parera has been very skillful when it comes to bowling and good batting.

[2] http://www.thesamaja.com/news_archive.php

Negative Sentences
These reflect negative opinion, emotion, feeling or sentiment such as expressions of judgement, negative attitude, criticism, failure, sadness, negative emotional state etc. Negative sentences tend to have negative affective words present in them. A few examples are listed below.

- ୬୯ ରନରେ ଦଳ ହସି ଓ ବିଜୟଙ୍କ ୱିକେଟ ହରାଇବା ପରେ ସୁରେଶ ରୈନା ୭ ରନରେ ଦର୍ଶକଙ୍କୁ ନିରାଶ କରିଥିଲେ
  Transliteration: 69 ranre dala hassi oh bijayanka wiket haraaibaa pare suresh raainaa 7 ranre dalanku niraasa karithile
  Affective Words: ହରାଇବା, ନିରାଶ
  English: After the team lost the wickets of both Hussey and Vijay at 69 runs, Suresh Raina had disappointed the team with 7 runs only.

- ଏହି ଅଞ୍ଚଳଟି ମାଓବାଦୀ ପ୍ରବଣ ହୋଇଥିବାରୁ ସବୁ ଯାତ୍ରୀଙ୍କ ଭୟ ହୋଇଥିଲା
  Transliteration: Ehi anchalati maaobaadi prabana hoithibaaru sabu jaatrinka madhyare bhaya houthilaa
  Affective Words: ପ୍ରବଣ, ଭୟ
  English: Travellers have been terrified/fearful because of the infestation of maoists in this area.

Neutral Sentences
Certain sentences may have neither positive nor negative opinion. These are labelled under the Neutral category. Very few cases may have both positive and negative sentiment where one does not necessarily dominate the other. Certain neutral sentences may not even have positive and negative phrases present in them. These typically lack affective words in them. Named entities which have positive or negative meaning in the language should not be considered as affective words as these don't contribute to the polarity of the sentence. Other than that, sentences which state a fact assuredly and which have an evidence to support the fact are also categorized under Neutral sentences. Factual statement occur regularly in news articles. These sentences express no feeling or emotion in them.
A few examples of neutral sentences are listed below.

- ପୁଲିସ ବର୍ତ୍ତମାନ ସୁଦ୍ଧା ଘଟଣାସ୍ଥଳରେ ପହଞ୍ଚିନାହିଁ
  Transliteration: Pulis barthamaan sudhaa ghatanaa sthalare pahanchinaahi
  Affective Words: None
  English: The police has not arrived at the scene yet.

- ଏହି ମ୍ୟାଚ୍ ସହ ଦୁବାଇରେ ଏହି ୨ ଦିନିଆ ସରିଜ୍ ମଧ୍ୟ ସମାପ୍ତ ହୋଇଛି
  Transliteration: Ehi myatch saha dubaire ehi 2 diniyaa siriz madhyan samaaptha hoichi
  Affective Words: None
  English: With this match, the two-day series

has ended/concluded in Dubai.

- ସ୍ୱାନ୍ ୩ଟି ଓ ଆଜ୍ମଲ ସେହ୍ଜାଦ୍ ୨ଟି ୱିକେଟ ପାଇଥିଲେ
Transliteration: Swaan 3ti oh ajmal sehjaad 2ti wiket paaithile
Affective Words: None
English: Swan and Ajmal Sehjhad got three and two wickets respectively.

- ଏଥିସହିତ ଏହି ଘଟଣାରେ ବ୍ୟବହୃତ ହିରୋହୁଣ୍ଡା ପ୍ୟାସନ ମୋଟରସାଇକେଲକୁ ମଧ୍ୟ ଜବତ କରାଯାଇଛି
Transliteration: Ethi sahitha ehi ghatanaare byabahrutha herohondaa passion motorcycle ku madhya jabath karaa jaayichi.
Affective Words: ଜବତ
English: Alongside this, the used hero-honda passion motorcycle has also been ceased.

- ଅଷ୍ଟ୍ରେଲିଆ ୬୫ରେ ଅଲ୍ ଆଉଟ ଭାରତ ୨୪ ୩ ରନ୍ ରେ ବିଜୟୀ
Transliteration:
Affective Words: ବିଜୟ
English: Australia : All-out in 65 runs. India win by 24 runs.

- ଏଥିରେ ମେୟର ଅନନ୍ତ ଜେନା ବିଏମସି କମିଶନର ଗଦାଧର ପରିଡ଼ାଙ୍କ ସମେତ ସଡ଼କ ଓ ପରିବହନ ବଭାଗର ସଚିବ ସତ୍ୟବ୍ରତ ସାହୁ ପୁଲିସ କମିଶନର ବିଜୟ କୁମାର ଶର୍ମା ଡିସିପି ହିମାଂଶୁ ଲାଲ ବିଧାୟକ ଅଶୋକ ପଣ୍ଡା ବିଜୟ ମହାନ୍ତି ଓ ବହୁ କର୍ପୋରରେଟର ଯୋଗଦେଇଥିଲେ
Transliteration: Ethire mayor annanth jenaa BMC commissioner gadhaadhar paridaanka sametha sadaka oh paribahana bibhaagara sachiba sathyabratha saahoo police commissioner bijaya kumaar sharmaa DCP himaanshu laal, bidhaayak ashok pandaa bijaya mahaanti oh bahu corporater jogaadeithile.
Affective Words: ଅନନ୍ତ, ବିଜୟ(Named Entities)
English: In this, Mayor Ananth Jena, BMC Commissioner Gadadhar Parida, Road and Vehicle Department's Satyabrata Sahoo, Police Commissioner Bijay Kumar Sharma, DCP Himanshu Lal, MLA Ashok Panda, Bijay Mahanti, and several other corporate officials had participated.

Other Guidelines
Each annotator was to annotate a given sentence as one of three labels: positive, negative or neutral. If the annotator was unsure about the label of a given sentence, they were advised to mark it as *unsure.* Some of these were later removed from the final dataset. The following were also discussed with the annotators in order to facilitate proper annotations:

- **Illocutionary Speech Acts** - Expression of intent via various speech acts contribute significantly in determining the sentiment associated with a sentence (Searle, 1975). Some of the positive speech acts include *motivating*, *praising*, *expression of gratitude*, *promising*, *congratulating*, *expression of admiration*, etc. Some of the negative speech acts include *expressions of judgement*, *criticism*, *convicting*, *banning*, *penalizing*, *regretting*, *disappointment*, etc. Annotators were made to understand and identify the usage of such acts in order to better identify the sentiment for such types of sentences.

- **Author's point of view** - It is important focus on the language used by the author of the article. The language used, gives insight on the point of view of the writer of the sentence. This further contributes to the sentiment associated with a given sentence (Lin et al., 2006). Understanding the sentence from the author's perspective based on the language (e.g - usage of affective words) used by the author should help determine the sentiment of a given sentence.

- **Annotator's point of view** - Every annotator has their own pre-conditioned biases associated with certain sentences based on social, cultural and economic conditions. For example, a sentence describing the English cricket team's victory over India could invoke negative sentiment, given the annotator's strong support of the latter. However, the actual label of such a statement would be positive because of the author's intention. Therefore, the annotators were advised to strictly avoid making annotation decisions based on their own point of view (e.g - personal prejudices) when it comes to such sentences. The annotators were instead recommended to focus on the author's perspective and language used by the authors when determining the label for such sentences.

3.2. Dataset Evaluation

A total of three Odia annotators were given the task to annotate the sentences into positive, negative and neutral classes. These annotators speak, read and write in Odia on a daily basis. If an annotator was unsure or confused about the label for a given sentence, they were advised to mark the same as *unsure.* For a given sentence, the final annotation label was determined by a majority rule. If two or more annotators marked a given sentence as *unsure*, then such sentences were removed from the dataset in order to avoid ambiguity. A total of 559 sentences were marked as positive. 574 sentences were marked as carrying negative sentiment. As expected, a good chunk of the sentences in the news dataset were neutral, numbering 912. 42 sentences had majority label of *unsure* and were removed from the dataset. Corpus statistics are reported in Table 1.
Fleiss' Kappa [3] Inter-annotator agreement score was calculated to help estimate the reliability of these an-

[3] `https://en.wikipedia.org/wiki/Fleiss'_kappa`

Data Type	#
Total Articles	175
Initial Sentences	2087
Positive Sentences	559
Negative Sentences	574
Neutral Sentences	912
Removed(unsure) Sentences	42
Token Count	29419
Final Sentences	2045
Inter-Annotator Agreement (**Fleiss Kappa**)	
$\kappa = \mathbf{0.791}$	

Table 1: Statistics for News-domain Dataset

notations. The formula for calculating Fleiss' Kappa is mentioned in the equation 1.

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \tag{1}$$

$\bar{P}$ is the sum of observed agreement and $\bar{P}_e$ is the sum of agreement by chance. We took a sample set of 550 sentences from the dataset in order to determine the Inter-annotator agreement score. An agreement score of $\kappa = \mathbf{0.791}$ is reported for the news domain dataset. This corresponds to substantial agreement..

4. Classification Experiments

In order to determine the baseline for the created dataset, we conducted a few experiments using Machine learning techniques. We employed useful features from previous experiments (Mohanty et al., 2018) along with the Odia sentiment lexicon. This helps provide an insight on the performance and reliability of the sentiment lexicon.

4.1. Experimental Setup

We divided the task of sentiment analysis into two separate classification problems. Firstly, we conducted Binary Classification with classifiers trained only on sentences labelled as positive and negative. Then we conducted ternary classification with the classifiers trained on the complete dataset including neutral sentences.

For each type of classification we first determined the baseline performance of various classifiers using a few of the best features from previous experiments. We then made use of Odia sentiment lexicon in order to identify affective words at sentence-level and add these as features for the classifiers. The feature using the lexicon contains two values: number of words referring to the positive sentiment and the number of words expressing negative sentiment. This feature was appended to the each TF-IDF vector representation for each sentence. 5 fold-cross validation was carried out on the dataset. Experiments were conducted using the scikit-learn (Pedregosa et al., 2011) library. We used four metrics to evaluate performance of various features and classifiers: **Precision**, **Recall**, **F1-Score**, and **Accuracy**.

4.2. Classifiers

We employed both **Support Vector Machines** and **Logistic Regression** as both performed reasonably well in baseline experiments. We also used **Random Forest** among the set of classifiers for the task sentence-level sentiment classification.

Random Forest serves as an ensemble of Decision Trees. Random forests construct multiple decision trees, considering the scores of each tree before deciding the final output. Unlike decision trees, Random forests reduce over-fitting due to inclusion of multiple trees.

4.3. Features

Based on their performance in previous (Mohanty et al., 2018) experiments, the following features were used to train the aforementioned classifiers:

- **TF-IDF Word-Level Features** - We incorporated both *unigrams* and *unigram-bigrams* for word-level features. Trigrams were not used because of the relatively smaller size of the dataset which, as a result, led to the presence of a large number of sparse trigrams. Trigrams should work more effectively on a much larger dataset where the sparsity of trigrams is reduced to a great extent.
- **TF-IDF Character-Level Features** - Character-Level TF-IDF features have proven to show consistent improvement over word-level features and are therefore used to train classifiers in these experiments. For baseline, we used *2-6 character n-grams* and *3-6 character n-grams* as features.
- **Affective Words from Odia Sentiment Lexicon** - The major objective of this chapter is to estimate the performance and reliability of the created Odia Sentiment Lexicon for the task of sentence-level sentiment classification. We captured positive and negative affective words at sentence-level and used them as added features to the classifiers. The results of the experiment report how effective these features were for the task at hand.

5. Binary Classification

The following are results of performance of various classifiers using different features for binary classification.

5.1. Baseline Feature-wise Performance

Word-Level Features

Table 3 shows the results of using TF-IDF Word-Level features for Binary Classification. Linear-SVM performs slightly better than Logistic Regression with an average accuracy of 78.2% for unigram and 80.2% for unigram-bigram features. Moreover the Precision, Recall and F1-Score for Linear-SVM using unigram-bigram features shows the best results among the rest of the models.

Character-Level Features
As expected, TF-IDF Character-Level features show consistent improvement in comparison to Word-Level features for all classifiers. Linear-SVM outperforms Logistic Regression and Random Forest consistently as can be observed from Table 4.

5.2. Using Odia Sentiment Lexicon

We identify affective positive and negative words present at sentence-level with the help the Odia sentiment lexicon. It is to be noted that 559 sentences are labelled as positive and 574 sentences are labelled as negative. Table 2 shows the coverage of words, from the sentiment lexicon, in the sentences of the dataset. Given the relatively small size of the dataset, this coverage should suffice for experiments.

Positive Sentences	559
Negative Sentences	574
Positive Words Found	333/1803
Negative Words Found	408/2846

Table 2: Binary Classification: Odia Sentiment Lexicon Coverage

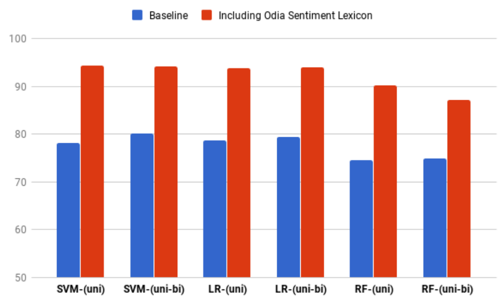

Figure 1: Accuracy improvements using Odia Sentiment Lexicon with Word-Level Features for Binary Classification

Inclusion of sentiment lexicon to word-level features has shown significant improvement in the performance of classifiers that can be observed in Table 5. Linear-SVM marginally outperforms Logistic Regression with an accuracy of 94.4%. Moreover, it can be observed that consistent performance improvements are seen for all the four three metrics of evaluation: Precision, Recall, F1-Score, and Accuracy. For example, Figure 1 helps comparing improvement in terms of accuracy when using the sentiment lexicon over baseline features. Similar improvements have been observed when using Odia sentiment lexicon with character-level TF-IDF features. Linear-SVM performs marginally better than Logistic Regression with the former having an accuracy of 95.2% and the latter having an accuracy of 94.4% as shown in Table 6. All four metrics show consistent improvement in performance when compared to baseline character-level features, across all classifiers. Figure 2 compares accuracy improvements between character-level baseline and the one including affective words from sentiment lexicon as features.

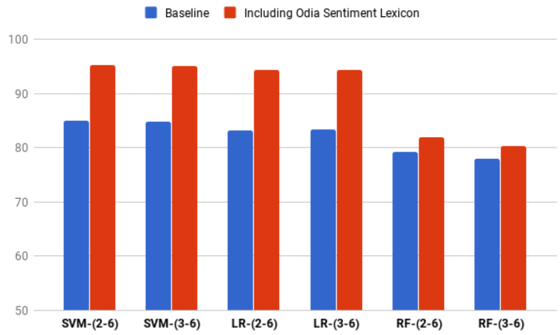

Figure 2: Accuracy improvements using Odia Sentiment Lexicon with Character-Level Features for Binary Classification

6. Ternary Classification

In case of Ternary Classification, sentences were classified into one of three categories: positive, negative or neutral. The following are results of performance of various classifiers using different features for ternary classification.

6.1. Baseline Feature-wise Performance

Word-Level Features
Table 8 shows the results of using TF-IDF Word-Level features for Ternary Classification. Logistic Regression performs slightly better than Linear-SVM with 57% accuracy. The former outperforms the latter in terms of Precision with Logistic Regression having a precision of 0.583 and Linear-SVM having a precision of 0.548. Linear-SVM performs marginally better than Logistic Regression when Recall and F1-Score are considered as metrics of evaluation. It is observed that Random Forest does not perform as well as the above two classifiers in case of Ternary Classification.

Character-Level Features
Linear-SVM outperforms Logistic Regression and Random Forest when using character-level features as observed in Table 9. Logistic Regression offers marginally better performance than Linear-SVM in terms of precision (Precision for LR is 0.675). However, the former fails to outperform the latter in the other three metrics of evaluation. Linear-SVM achieves highest accuracy of 62.8% for Ternary Classification using character-level TF-IDF features.

6.2. Using Sentiment lexicon

For ternary classification, the coverage of the Odia sentiment lexicon was measured and the results of the same are shown in Table 7.
Comparing with coverage in Table 2 it is clear that a few positive and negative words have also been found among neutral sentences. It is observable that, even after having a large number of neutral sentences (44.5%

Model	Features	Precision	Recall	F1-Score	Accuracy(%)
Linear SVM	uni	0.783	0.782	0.781	78.2
	uni-bi	**0.803**	**0.802**	**0.802**	**80.2**
Logistic Regression	uni	0.787	0.786	0.786	78.7
	uni-bi	0.795	0.794	0.794	79.4
Random Forest	uni	0.751	0.745	0.743	74.6
	uni-bi	0.754	0.748	0.747	74.9

Table 3: Binary Classification: Using Only Word-Level TF-IDF Features

Model	Features	Precision	Recall	F1-Score	Accuracy(%)
Linear SVM	**(2-6)-gram**	**0.850**	**0.849**	**0.849**	**84.9**
	(3-6)-gram	0.849	0.848	0.848	84.8
Logistic Regression	(2-6)-gram	0.832	0.831	0.831	83.1
	(3-6)-gram	0.835	0.833	0.833	83.4
Random Forest	(2-6)-gram	0.798	0.790	0.789	79.2
	(3-6)-gram	0.784	0.779	0.778	78.0

Table 4: Binary Classification: Using Only Character-Level TF-IDF Features

Model	Features	Precision	Recall	F1-Score	Accuracy(%)
Linear SVM	**uni**	**0.944**	**0.943**	**0.943**	**94.4**
	uni-bi	0.941	0.940	0.940	94.1
Logistic Regression	uni	0.939	0.938	0.938	93.8
	uni-bi	0.940	0.939	0.939	94.0
Random Forest	uni	0.902	0.900	0.901	90.1
	uni-bi	0.874	0.871	0.872	87.2

Table 5: Binary Classification: Using Word-Level TF-IDF Features with Sentiment Lexicon

Model	Features	Precision	Recall	F1-Score	Accuracy(%)
Linear SVM	**(2-6)-gram**	**0.952**	**0.952**	**0.952**	**95.2**
	(3-6)-gram	0.951	0.951	0.951	95.1
Logistic Regression	(2-6)-gram	0.944	0.943	0.943	94.4
	(3-6)-gram	0.945	0.944	0.944	94.4
Random Forest	(2-6)-gram	0.822	0.818	0.817	81.9
	(3-6)-gram	0.805	0.802	0.802	80.3

Table 6: Binary Classification: Using Character-Level TF-IDF Features with Sentiment Lexicon

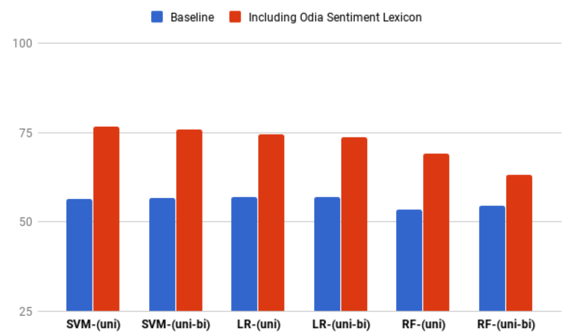

Figure 3: Accuracy improvements using Odia Sentiment Lexicon with Word-Level Features for Ternary Classification

Positive Sentences	559
Negative Sentences	574
Neutral Sentences	912
Neutral (No Affective)	417
Neutral (POS = NEG)	114
Neutral (POS > NEG > 0)	241
Neutral (NEG > POS > 0)	140
Positive Words Found	367/1803
Negative Words Found	446/2846

Table 7: Ternary Classification: Odia Sentiment Lexicon Coverage

of dataset), coverage has not increased significantly. This is due to lack of such affective words in factual statements. We have also observed that about 60% of the neutral sentences either contain no affective words or same number positive and negative words. We identify affective words in sentences and add them as features in addition to the baseline features, in order to help in Ternary classification.

In Table 10, it is observed that usage of sentiment lex-

Model	Features	Precision	Recall	F1-Score	Accuracy(%)
Linear	uni	0.545	0.517	0.519	56.5
SVM	**uni-bi**	0.548	**0.519**	**0.522**	56.6
Logistic	uni	0.567	0.513	0.515	56.9
Regression	uni-bi	**0.583**	0.505	0.507	**57.0**
Random	uni	0.509	0.489	0.487	53.4
Forest	uni-bi	0.515	0.484	0.483	54.5

Table 8: Ternary Classification: Using Only Word-Level TF-IDF Features

Model	Features	Precision	Recall	F1-Score	Accuracy(%)
Linear	(2-6)-gram	0.625	0.562	0.571	62.1
SVM	**(3-6)-gram**	0.640	**0.567**	**0.576**	**62.8**
Logistic	(2-6)-gram	0.655	0.531	0.536	60.6
Regression	(3-6)-gram	**0.675**	0.523	0.527	59.9
Random	(2-6)-gram	0.556	0.522	0.523	55.8
Forest	(3-6)-gram	0.539	0.515	0.516	54.3

Table 9: Ternary Classification: Using Only Character-Level TF-IDF Features

Model	Features	Precision	Recall	F1-Score	Accuracy(%)
Linear	**uni**	**0.767**	**0.751**	**0.756**	**76.7**
SVM	uni-bi	0.756	0.737	0.743	75.7
Logistic	uni	0.749	0.724	0.730	74.5
Regression	uni-bi	0.745	0.715	0.723	73.7
Random	uni	0.686	0.683	0.675	69.2
Forest	uni-bi	0.619	0.593	0.597	63.2

Table 10: Ternary Classification: Using Word-Level TF-IDF Features with Sentiment Lexicon

Model	Features	Precision	Recall	F1-Score	Accuracy(%)
Linear	**(2-6)-gram**	**0.780**	**0.753**	**0.762**	**77.6**
SVM	(3-6)-gram	0.778	0.753	0.761	77.4
Logistic	(2-6)-gram	0.756	0.725	0.733	74.7
Regression	(3-6)-gram	0.754	0.725	0.733	74.6
Random	(2-6)-gram	0.553	0.521	0.524	57.2
Forest	(3-6)-gram	0.548	0.524	0.525	56.2

Table 11: Ternary Classification: Using Character-Level TF-IDF Features with Sentiment Lexicon

icon does show consistent improvements for all word-level features across all three classifiers. Linear-SVM beats Logistic Regression and Random Forest consistently across all four metrics of evaluation. Precision, Recall and F1-Scores for Linear-SVM are highest with values of 0.767, 0.751 and 0.756 respectively. The highest accuracy for Linear-SVM is 76.7% followed by Logistic Regression with 74.5% accuracy. When comparing these with baseline word-level features for Ternary classification (Figure 3), the reliability of Odia sentiment lexicon can be deduced. For example, the figure shows a maximum improvement in accuracy of 20% for Linear-SVM with unigram features upon inclusion of identified affective words as features.

Similarly, inclusion of identified affective words as features along with character level features shows consistent improvements for Linear-SVM and Logistic Regression (Figure 4). The former comes on top with maximum accuracy of 77.6% whereas the latter shows a comparable accuracy of 74.7%. As observed in Table 11, in other three evaluation metrics, Linear-SVM consistently outperforms other classifiers with highest metric values of 0.78, 0.753 and 0.762 for Precision, Recall and F1-Score, respectively. Random Forest barely shows any increase in accuracy upon usage of sentiment lexicon with character-level features as can be seen in Figure 4.

7. Conclusion

This paper describes the creation of an annotated corpus of 2045 Odia sentences from articles in news domain. We discussed the annotation guidelines used to annotate these sentences into three categories: positive, negative and neutral. A substantial inter-annotator agreement score of **0.791** was obtained for the dataset. We performed baseline experiments using standard word and character-level features and machine learning techniques in order to conduct both binary and ternary sentiment classification. One major

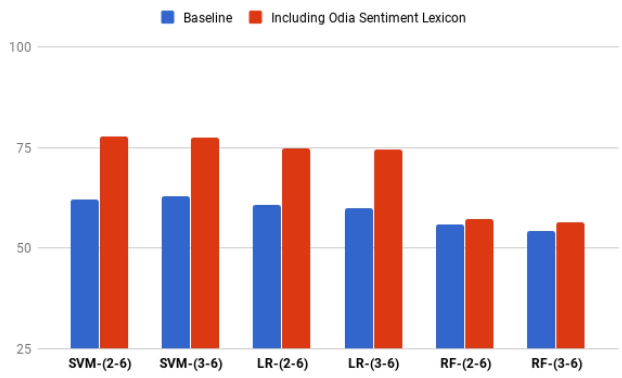

Figure 4: Accuracy improvements using Odia Sentiment Lexicon with Character-Level Features for Ternary Classification

objective of this chapter was to test the performance of Odia sentiment lexicon. We included identified positive/negative words, present at sentence-level, as features to various classifiers. It was observed that usage of Odia sentiment lexicon showed consistent and significant improvements in the overall performance of classifiers for both binary and ternary sentiment classification. This testifies the reliability of Odia sentiment lexicon for sentiment analysis related tasks. As an extension to this work, we would like to leverage word vector representations for Odia language and hence create better sentiment analysis models.

8. Bibliographical References

Bhattacharyya, P. (2010). Indowordnet. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10).*

Das, A. and Bandyopadhyay, S. (2010). Sentiwordnet for indian languages. In *Proceedings of the Eighth Workshop on Asian Language Resouces*, pages 56–63, Beijing, China, August. Coling 2010 Organizing Committee.

Lin, W.-H., Wilson, T., Wiebe, J., and Hauptmann, A. G. (2006). Which side are you on? identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 109–116.

Mohanty, G., Kannan, A., and Mamidi, R. (2017). Building a SentiWordNet for Odia. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 143–148, Copenhagen, Denmark, September. Association for Computational Linguistics.

Mohanty, G., Mishra, P., and Mamidi, R. (2018). Sad or glad?: Corpus creation for odia poetry with sentiment polarity information.

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., and Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830.

Searle, J. R. (1975). A taxonomy of illocutionary acts.