A Qualitative and Quantitative Analysis of the Political Discourse on Nepalese Social Media

Mandar Sharma [a], Dinesh Baniya Kshatri [b]

[a, b] *Department of Electronics and Computer Engineering, Pulchowk Campus, IOE, TU Nepal*
Corresponding Email: [a] mandar.sharma@ioe.edu.np

Abstract

The political discourse that takes place in the Nepalese online communities has been an uncharted territory for Opinion Mining. This research attempts to analyze the feasibility of implementation of Computational Social Science in the public sphere of Nepalese social media using two case studies. The first analysis was of a minor political controversy that was created by the speech given by Prime Minister Sher Bahadur Deuba at the World Leaders Forum organized at Columbia University [1] (Case I). The speech was considered to be vacuous and misrepresentative of Nepalese sentiment by a majority of the Nepali population. The second analysis was of a major political controversy caused by the recent alliances by major political parties in Nepal, leading to the formation of the Democratic Alliance and the Left Alliance [2] (Case II). The alliance was made in order to have a stronger hand in the upcoming provincial and parliamentary elections. In both cases, the qualitative and quantitative involvement of the Nepalese people in social-media based political discourse was analyzed.

Keywords

Computational Social Science – Opinion Mining

1. Introduction

Although Opining Mining in Social Media [3], or targeted mining for political discourse [4], have been popular areas for research, an analysis of political discourse or any kind of discourse on Nepalese social media has been rudimentary at best [5] [6]. This research explores the feasibility of applications of Computational Social Science in the Nepalese online community by presumption of two measures: The quantity of public responses with respect to the topics of the case study and a qualitative analysis that represents the sentiment of the said responses.

Nepal has a current population of 29,187,037 and according to the June 2017 Facebook statistics [7], the use of social media has penetrated to 21.9% of the population, resulting in 6,400,000 Nepalese Facebook users as of 20217. It has been estimated that approximately one-third of social media users partake in expressing their political opinions, while only 9% of the users report being actively involved in political discourse [8], leading to an assumed population of 576,000 users who actively partake in public political

discourse. As this study draws data from the public sphere of social media, the commenters and involved users fall under the category of active involvement.

Responses from the Facebook pages and YouTube channels of the most popular news and media outlets in Nepal have been scraped, resulting in a total response set of 11,737 responses (consisting of 3,143 textual responses) for the first case of a minor political controversy and a total response set of 268,989 responses (consisting of 5,762 textual responses) for the second case of a major political controversy [9]. This indicates that although the Nepalese social media does not provide adequate data source for an analysis of minor political controversies (2.03% penetration), it does provide a promising source for analysis of major socio-political controversies (46.69% penetration).

2. Data Acquisition

This data source was scraped from Facebook pages of popular media outlets using Facebook's Graph API (Application Programming Interface) and from

Table 1: Prime Minister Deuba's Speech at the World Leaders Forum (Case I)

Source	Non Textual Response	Textual Response
Men's Room Reloaded	7,945	2,545
Habamoment	290	104
Voice of Nepal	253	353
PM Deuba's Page	183	64
Youtube (Collective)	N/A	77

Table 2: Democratic and Leftist Alliance (Case II)

Source	Non Textual Response	Textual Response
News24 Nepal	64,187	1,021
Ekantipur	31,225	764
Nagarik News	47,037	746
Radio Kantipur	30,895	670
Taja Khabar	36,316	650
Light Nepal	12,304	491
DC Nepal	13,985	486
BBC Nepal	20,006	484
Avenues News	1,952	27
Youtube (Collective)	N/A	420

YouTube comments of the videos by the said media outlets[9]. Personal comments/responses from individuals are referred to as 'textual responses', while reactions (like, happy, sad or angry) to public posts are referred to as 'non textual response'.

As the language used for the majority of the communications is the Devanagari script, Google's translation service was used to convert those responses to English for effective classification for sentiment analysis. Although the conversion leads to inaccurate grammar in the destination language (English), the sentiments presented are substantially retained.

The acquired dataset shows that although the penetration for minor controversies is 2.03%, the Nepalese online community is a substantial ground for Opinion Mining for major socio-political issues as data from this research provides a penetration of 46.69%.

3. Sentiment Analysis

Naive Bayes is the family of probabilistic supervised learning algorithms that are based on Bayes' theorem, but with the "naive" assumption that every pair of features are independent of each other. A Naive Bayesian Classifier was used for sentiment analysis of the obtained responses, as it has been estimated to have a classification accuracy of at least 80%[10].

The probability that a document with feature vectors x_1 through x_n be classified as class C is given by Naive Bayes' theorem as:

$$P(C|x_1,...,x_n) \propto P(C) \prod_{i=1}^{n} P(x_i|C)$$

3.1 Case I : The controversy regarding Prime Minister Deuba's Speech (Minor Controversy)

In an analysis of 2,182 textual responses (as 961 of the remaining responses were incomprehensible), it was
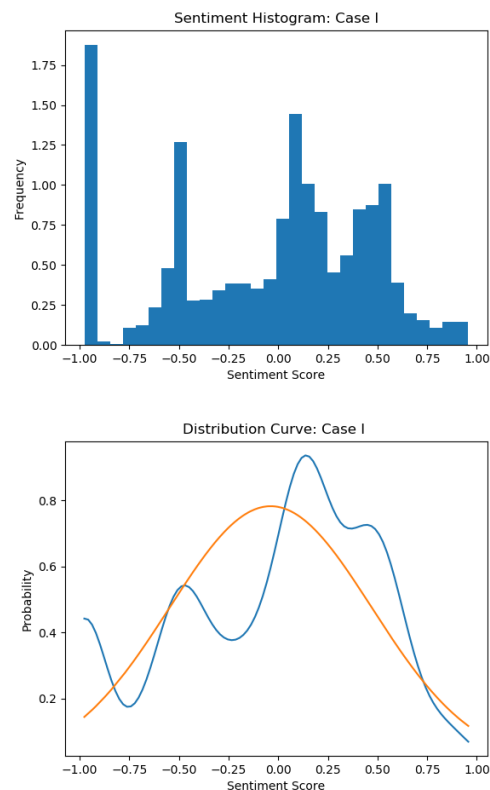

Figure 1: Histogram and KDE(Blue)/PDF(Orange) Curve for Case I (Minor Controversy)

found that 1,276 responses (58.47%) were labelled as positive where as 897 responses (41.10%) were labelled as negative. The four observable peaks in the sentiment histogram (Figure 1) show that the highest frequency of responses constitute of ones that are highly negative, mildly negative, almost neutral or sarcastically positive. The mean of the dataset was found to be -0.0380 (μ) with a standard deviation of 0.5094 (σ). In a sample of 300 responses, it was found that 22.38% of the sarcastic comments were labelled positive, leading to the disagreement of the positive spectrum of the histogram to the presumed sentiment of the Nepalese people. The negative classifications, however, are highly accurate, as demonstrated with the following sample cases. This data demonstrates that the majority of the Nepalese commenters either responded to the controversy with a sarcastic remark or a strong negative remark.

Sets of randomized sample responses (table 3) were pitted against human curated results for predicting accuracy of the classifier.
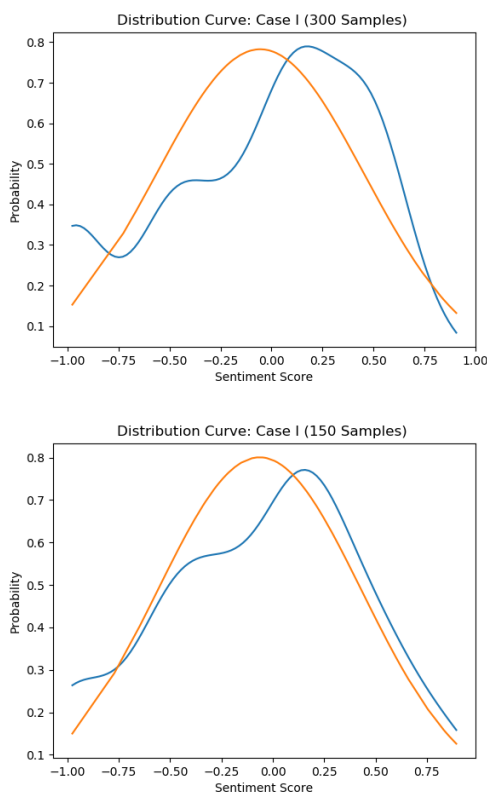
Sarcastic Positive Responses: 30 out of 143
True Positive (TP) Rate: 45.52%
False Positive (FP) Rate: 32.08%
Sarcastic responses classified as Positive: 22.38%
True Negative (TN) Rate: 96.61%
False Negative (FN) Rate: 3.38%

From the statistical analysis of the sample, it can observed that the true positive rate has suffered due to sarcastic responses which make up 22.38% of the total positive responses. However, it can also be seen that the classifier predicts negative responses with a high degree of accuracy (96.61%). The sample mean for the 300 randomized responses is -0.0553 (X_1) and the standard deviation is 0.5097 (s_1), while the sample mean for the 150 randomized responses is -0.0646 (X_2) and the standard deviation is 0.4981 (s_2), substantiating the accuracy of sample sets.

3.2 Case II : The controversy regarding the Democratic and Leftist Alliance (Major Controversy)

In an analysis of 5,526 textual responses (as 236 of the remaining responses were incomprehensible), it was found that 3,239 responses (58.61%) were labelled as positive where as 2,089 responses (37.80%) were labelled as negative. The histogram (figure 3) suggests that although the opinion among Nepalese people is distributed between the extremities of positive and negative sentiment scores, a majority of the people hold a slightly positive opinion towards the formed alliances, as indicated by the peak in the histogram. The mean of the dataset was found to be 0.1171 (μ) with a standard deviation of 0.4406 (σ). Also, the positive bias due to sarcastic responses was not prominent in this case, thus yielding a greater positive classification accuracy. Although, we do need to keep in mind the a false positive rate of 35.18% while drawing conclusions from the positive spectrum of this histogram.

Sets of randomized sample responses (table 4) were

Figure 2: KDE(Blue)/PDF(Orange) Curve for Case I (Sample Population)

Table 3: Confusion Matrix for Case I (300 Samples)

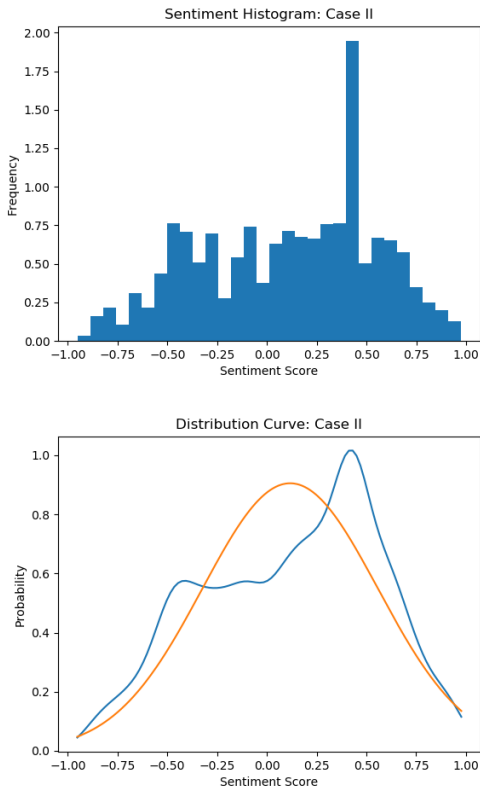| N = 300 Samples	Predicted: Positive (134)	Predicted: Negative (59)
Actual: Positive (61)	TP = 61	FN = 2
Actual: Negative (57)	FP = 43	TN = 57

Figure 3: Histogram and KDE(Blue)/PDF(Orange) Curve for Case II (Major Controversy)
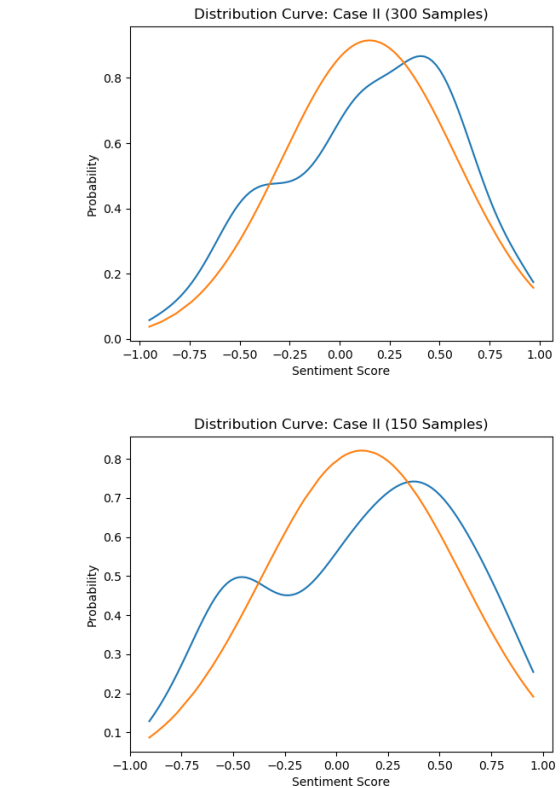
Figure 4: KDE(Blue)/PDF(Orange) Curve for Case II (Sample Population)

Table 4: Confusion Matrix for Case II (300 Samples)

N = 300 Samples (Random)	Predicted: Positive (108)	Predicted: Negative (62)
Actual: Positive (70)	TP = 70	FN = 4
Actual: Negative (58)	FP = 38	TN = 58

pitted against human curated results for predicting accuracy of the classifier.

True Positive (TP) Rate: 64.81%
False Positive (FP) Rate: 35.18%
True Negative (TN) Rate: 93.54%
False Negative (FN) Rate: 6.45%

From these statistics, it can be observed that the absence of sarcastic responses has significantly increased the positive classification accuracy from 45.52% (previous case) to 64.81%. The sample mean for the 300 randomized responses is 0.1494 (X_1) and the standard deviation is 0.4359 (s_1), while the sample mean for the 150 randomized responses is 0.1243 (X_2) and the standard deviation is 0.4855 (s_2), substantiating the accuracy of sample sets.

4. Discussion and Conclusion

The study offers three conclusive results. Primarily, it provides reasoning that although the Nepalese social media has not bloomed substantially for mining sentiments with respect to minor controversies (Case I, PM Deuba's Speech), it does provide substantial grounds for analysis of major socio-political issues (Case II, Political Alliance), providing a penetration of 46.69% to the representative population. Second, it shows that the controversy regarding PM Deuba's speech at the World Leaders Forum was met with a majority of extremely negative or sarcastic responses from the Nepalese people. Third, it shows that the major political movement in the recent Nepalese scenario, the formation of Democratic and Leftist Alliance, has led to sentiments among the Nepalese people that is distributed between the extremities of positive and negative sentiment scores, a majority of the people hold

a slightly positive opinion towards the formed alliances.

5. Future Work

The first claim of this study, presenting the Nepalese online community as a viable data source for opinion mining with regards to major socio-political changes, opens doors for substantial future research. A preliminary analysis for upcoming national elections based on candidate popularity in social media, targeted disaster relief for earthquakes based on data from social media, an analysis of the mental health of Nepalese citizens based on their content posted on social media, are among the many potential areas of expansion of this research. As stated above, till now, the Nepalese online community has been an uncharted territory for Computational Social Science.

References

[1] Sher Bahadur Deuba, Prime Minister of Nepal, at the Columbia University World Leaders Forum. https://www.youtube.com/watch?v=VxP_GdHfguA, 2017. [Online].

[2] Democratic Alliance: Nepali Congress, Nepal Loktantrik Forum, Rastriya Prajatantra Party and Naya Shakti Party. Left Alliance: CPN UML, CPN Maoist and majority of the communist parties. https://goo.gl/ZRzjQp, 2017. [Online].

[3] Bo Pang and Lee Lillian. Opinion mining and sentiment analysis. 2008.

[4] Pawel Sobkowicz, Michael Kaschesky, and Guillaume Bouchard. Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. 2012.

[5] Ujjwal Acharya. Online media in nepal: Need for policy intervention. 2012.

[6] Ujjwal Acharya, Tilak Pathak, Bhuwan KC, Bikash Karki, Rajneesh Bhandari, Umesh Shrestha, and Rishikesh Dahal. Journalist & social media 2011 national survey on nepali journalists. 2011.

[7] Social media penetration in Nepal. https://goo.gl/CnUKLN, 2017. [Online].

[8] Maeve Duggan and Aaron Smith. The political environment on social media. 2016.

[9] Research Data Set. https://goo.gl/1hXF79, 2017. [Online].

[10] Lopamudra Dey, Sanjay Chakraborty, Anuraag Biswas, Beepa Bose, and Sweta Tiwari. Sentiment analysis of review datasets using naïve bayes and k-nn classifier. 2016.