
Three years – three moves – government verdicts to renovate customary ‘Bharat’ to contemporary ‘India’: evaluation of opinions from citizens

Riktेश Srivastava

Skyline University College,
Sharjah, UAE
Email: rsrivastava@skylineuniversity.ac.ae

Abstract: Data is an important asset for the government and enormous data are collected every day. Assortment of data is one trait, deriving meaningful insights from these data is important too. Emergence of ‘data analytics’ creates link between effective governance and improved service delivery. Two types of data collected by the government are structured and unstructured. Structured data are collected through operational work and are used for decision making since long. Unstructured data now forms an integral form of decision making, are the ones collected from social media, websites and other online media. Government of India (GOI) identified the need of integrating unstructured data (especially social media). The study collects opinions concerning three historic moves – digital India, demonetisation, and GST, through eight online media including six Hindi news channels and news two English newspapers. The texts are evaluated based on ‘profile of mood states’ (POMS) tagging and identifies opinions.

Keywords: digital India; demonetisation; GST; profile of mode states; opinion analysis; Bayes algorithm; India.

Reference to this paper should be made as follows: Srivastava, R. (2019) ‘Three years – three moves – government verdicts to renovate customary ‘Bharat’ to contemporary ‘India’: evaluation of opinions from citizens’, *Int. J. Business and Data Analytics*, Vol. 1, No. 1, pp.54–68.

Biographical notes: Riktेश Srivastava received his PhD in Electronics and Management qualification from the IIMA, Certification in Marketing from the Wharton School and Electronic Commerce from the NTU, Singapore. Currently, he is an Associate Professor (Information Systems) at the Skyline University College, Sharjah, UAE. His areas of interests are business analytics, queuing theory and electronic commerce.

1 Introduction

It is evident that spread of news online empowers the government to get people feedback to form communication strategies via co-creation (Füller, 2010), participation in real-time exchanges (Yan, 2011), improve strategy using a low-cost multidimensional information channel (Sinickas, 2011) and audience extension (Booth, 2011). These political opinion analysis using online media (both social media and online published news) has attracted

many researchers to dig deep into data collection and analysis (Kermanidis and Maragoudakis, 2013). Such opinion analysis also comprehends to business sustainability (Eryürük et al., 2014) and GOI used both types of news to twig an broad selection of interactional tools (texts and videos) and to create a two-way dialogue opportunities. Though GOI is quite aggressive in using online platform as a tool of communication and identify how communities can work together to address social and economic challenges, but it failed to identify the impact of these platforms. Twitter is one of the most effective tool for collecting responses from consumers in business environment (Klotz et al., 2016) and GOI trust on online medium is so intense, that it partnered with Twitter to deliver tweets as SMS on mobile phones (HT, 2015). Although Twitter helped GOI to send the messages to public and reach more than 700 million mobile phone users, apart from estimated 33 million Twitter population in India, however, opinion analysis of short texts is challenging because of the restricted circumstantial evidence. YouTube videos usage in India is immense with 70% of its users are below the age of 35, 1 in 5 users creates its own video, 3 in 5 reads the comments on video and 7 in 10 reads those comments, with total of YouTube mobile viewership hitting 180 million (Menon, 2017). Apart from using online media to post the updates, citizen opinions also get affected by the news published in newspapers. The rise of media in India is obsessed by swift revolution of muddling new technologies with media, and this combination has created an empowered and aggressive media (Antilla, 2014). Both news channels and newspapers in India started to use online mediums as a superfluous source to publish news and reach broader spectators. GOI though used TV and newspapers to apprise Indian citizens of the three historic movements; it stretched a wider audience via volatile use of online media by these two vehicles.

Notwithstanding the multiplicity of uses evolving from such a simple communication channel, it has been noted that posts on online media tend to fall in one of two types (Bollen et al., 2009): first to post the news, and, secondly to use them to share information (Naaman et al., 2010). In both cases, the responses from citizens deliver evidence about the mood states. In both the cases, mood expressions can be captured through the opinions of citizens, which are regarded as microscopic instantiations of mood. It was also evident that the opinions generated over a given time also reveals variations in the state of public mood at a grander scale.

The existing study amasses the texts based on news from Hindi news channels (through YouTube, based on news from Hindi news channels) and traditional media (grounded on news from two leading English newspapers from India). Hindi news channels selected for the study are Aaj Tak, ABP News, CNBC, IBN7, NDTV and Zee news, and the two English newspapers included are Times of India, and The Indian Express. The responses from public for Digital India were from July 01, 2015 to March, 31, 2018 summing to 1,107 responses, for demonetisation, the responses were collected from November, 09 2016 to March, 31 2018 with 2,844 responses, and for GST, there were 2,641 responses from July 1, 2015 to March, 31 2018. These responses were then investigated for opinion mining. Opinion mining is an operative methodology to evaluate citizens feedback and comment (Al-Ayyoub et al., 2015). POMS tagging was adopted to identify the opinions and comprises of seven types of mood domains {*fatigue, anger, vigour, confusion, depression, tension and friendliness*} (Shahid et al., 2011). Six of mood states are used in the research, spawned via the analysis of such combined mood aggregators. The advantage of using publicly accessible online data for such historic

movements through opinions collected online pointedly reduces the costs, efforts and time.

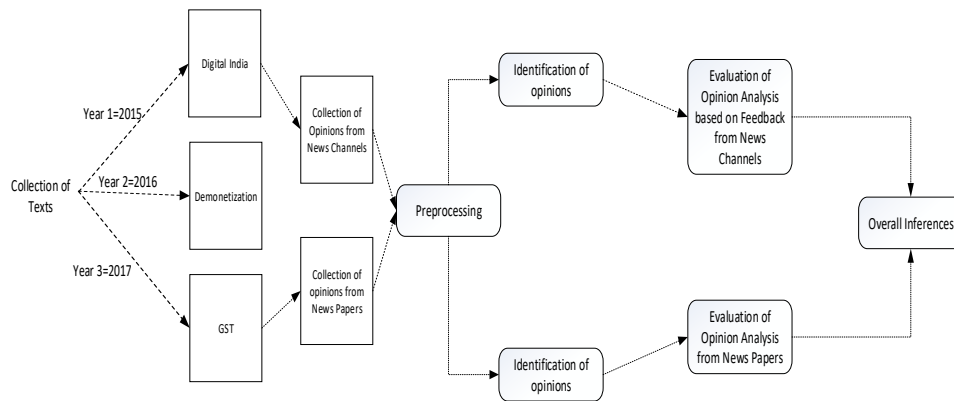
The paper is distributed into five sections: Section 2 expounds the framework supported for the study.

Section 3 explicates the code snippet adopted for the study, explains the preprocessing steps and enumerates the mathematical evaluations for POMS analysis technique. Section 4 elucidates the research analysis, where, the outcomes of the experiments are portrayed for all eight online media. Section 5 pulls the conclusion and inferences of the outcomes.

2 Framework for the study

Figure 1 elaborates the online media analysis framework, for both news channels and newspaper respectively.

Figure 1 Proposed online media analysis framework



As mentioned in Figure 1, the framework is divided into three parts. Part 1 collects the texts from varied sources. Online texts contain usually lots of noise and uninformative parts (Haddi et al., 2013), and requires cleansing through preprocessing (Malley et al., 2016), used in part 2 of the framework via five steps. Part 3 uses POMS tagging to identify the opinions of citizens. The outcome of POMS tagging is in matrix format and needs to be converted into vector format for detailed analysis (Sreeja, 2016). The conversion was performed using Bayes theorem and results are depicted as overall inferences part of overall framework.

3 Explanation of parts for proposed framework

As mentioned in Section 2, the three parts of proposed framework are

- part 1: collection of texts
- part 2: preprocessing of texts

- part 3: POMS tagging for opinion analysis.

This section clarifies each of these parts in facet.

3.1 Collection of texts

As mentioned in Figure 1, there are two sources for text collection, one from YouTube and secondly from newspapers.

The text collection from six Hindi news channels, through YouTube are collected using R programming using the following code snippet.

```
library(SocialMediaLab)
apikey<- "xxxxxxxxXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
key <- AuthenticateWithYoutubeAPI(apikey)
video <- c('videoIDs')
ytdata<- CollectDataYoutube(video, key, writeToFile = FALSE)
write.csv(ytdata, "abc.csv", row.names = F)
```

Similarly, the opinions of public from two English newspapers are collected using an API used for R programming language. The code snippet is given as under:

```
library(quantmod)
library(XML)
rawPMI<- readHTMLTable('http://www.timesofindia.com/GST/content.html?Item
Number=10752')
PMI <- data.frame(rawPMI[[1]])
```

3.2 Preprocessing steps

The step for preprocessing is to eradicate the words (not texts, as witnessed that some texts includes mishmash of English and other language words), not in English. The other five steps in preprocessing are explicated in step wise means in this part.

Step 1 Transformation

The text collected are always in unstructured and raw format. Transformation converts this unstructured raw format into more meaningful format for analysis. The steps for transformation are:

- a converts all text into lowercase
- b remove all diacritics or accents in text, as héllö→ hello
- c most of users give links online. These texts acts as a single word and results in wrong identification of either sentiments or topics. <a href> i love my country gets converted into → i love my country
- d removes all URLs from texts.

The steps can be represented mathematically by equation (1) as:

$$\sum_{i=1}^n \sum_{j=1}^m t_{ij} = \sum_{i=1}^n \prod_{j=1}^m \begin{bmatrix} (t_{11}) \\ (t_{12}) \\ \vdots \\ (t_{nm}) \end{bmatrix} = \sum_{i=1}^n \prod_{j=1}^m \begin{bmatrix} (Word_{11} Word_{12} \dots Word_{1j}) \\ (Word_{21} Word_{22} \dots Word_{2j}) \\ \vdots \\ (Word_{n1} Word_{n2} \dots Word_{nm}) \end{bmatrix} \quad (1)$$

where i is number of texts where each texts contain j words. Notice from equation (1) that each word of texts are not disjointed and are embodied as sentences.

Step 2 Tokenisation

The texts generated from step 1 in its naïve form is not suitable for processing, where all possible documents are available in digital formats, without any specific meaning (Subitha and Sujatha, 2017). Tokenisation is a process of breaking texts into words [as stated in equation (1)]. These words are smaller components called tokens. To do so, broad rules of splitting each word in the text by:

- 1 Punctuations are also considered as tokens separately from words. These tokens are represented by either white space or full stop.
- 2 Emoticons are also used as tokens. The impact of emoticons is relatively substantial in categorising sentiment analysis.

Equation (1) can now be redefined as in equation (2) as:

$$= \sum_{i=1}^n \prod_{j=1}^m \begin{bmatrix} (Word_{11})(Word_{12}) \dots (Word_{1j}) \\ (Word_{21})(Word_{22}) \dots (Word_{2j}) \\ \vdots \\ (Word_{n1})(Word_{n2}) \dots (Word_{nm}) \end{bmatrix} \quad (2)$$

Notice, that each texts are now separated into expressive tokens.

Step 3 Normalisation

Step 3 receipts all words as cited in equation (2) as input and executes stemming and lemmatisation. Stemming is heuristic progression of axing of derivational affixes. There are likelihoods that after stemming, some words may appear irrelevant, though relevant (AleEbrahim and Fathian, 2013). Lemmatisation does this task by the using vocabulary and morphological analysis of words (Manning et al., 2008). Porter 2 Stemmer also called Snowball stemmer (Porter, 1997) is used for this step. The words are muddled based on affixes and stemmed and lemmatised consequently. The process is explained in equation (3):

$$= \sum_{i=1}^n \prod_{j=1}^m \left[\begin{array}{c} (Word_{11})(Word_{12}) \dots (Word_{1j}) \\ (Word_{21})(Word_{22}) \dots (Word_{2j}) \\ \vdots \\ \vdots \\ \vdots \\ (Word_{n1})(Word_{n1}) \dots (Word_{nm}) \end{array} \right] = \left[\begin{array}{c} (w_{11}, w_{12}, \dots, w_{1m}) \leftarrow x_1 \\ (w_{21}, w_{22}, \dots, w_{2m}) \leftarrow x_2 \\ \vdots \\ \vdots \\ (w_{n1}, w_{n2}, \dots, w_{nm}) \leftarrow x_n \\ v_1, v_2, \dots, v_m \end{array} \right] \quad (3)$$

where $w_{11}, w_{12}, \dots, w_{1n}$ are words with similar affixes and mapped to x_1 , $w_{21}, w_{22}, \dots, w_{2n}$ are mapped to x_2 and words v_1, v_2, \dots, v_m does not have affirmative affixes and thus remain invariant.

The equation can be rewritten as equation (4):

$$= \left[\begin{array}{c} (ws_{11}, ws_{12}, \dots, ws_{1m}) \\ (ws_{21}, ws_{22}, \dots, ws_{2m}) \\ \vdots \\ \vdots \\ \vdots \\ (ws_{n1}, ws_{n2}, \dots, ws_{nm}) \end{array} \right] \quad (4)$$

Step 4 Filtering

Filtering involves eliminating common words, also called stopwords, from equation (4). The stopwords used was proposed by Liu and Hu (Liu et al., 2005). The number of words substantially reduces from equation (4) after stop words, as shown in equation (5)

$$= \sum_{i=1}^n \prod_{j=1}^k \left[\begin{array}{c} (ws_{11}, ws_{12}, \dots, ws_{1k}) \\ (ws_{21}, ws_{22}, \dots, ws_{2k}) \\ \vdots \\ \vdots \\ \vdots \\ (ws_{n1}, ws_{n2}, \dots, ws_{nk}) \end{array} \right] \quad (5)$$

Notice in equation (5) that amount of words are substantially concentrated from k to m , where

$$k \ll m.$$

Step 5 ngram

Next step is termed as ngram, which is sequence of n words, is used for filtration for key-phrase extraction. Key-phrase extraction uses a combination of statistical feature (based on weighted betweenness centrality scores of words) and co-location strength (based on nearest neighbour) (Kumar et al., 2016). The

result of the analysis generates combination of words are principally a set of co-occurring word, and, while computing the ngram typically move one word forward (Ganesan et al., 2012). The generic representation of ngram is:

$$ngram_t = k - (t - 1) \quad (6)$$

where

k = numbers of words in a text t , then

Bigrams (two-gram) are used for the research as it includes features comprising of sets of two adjacent words from text. It was observed that unigram cannot capture phrases and multi-word expressions, effectively ignoring any word order dependence (Tan et al., 2002). As an example, words like ‘not excited’, ‘not satisfied’ clearly say that the sentiment is negative, but a unigram might fail to identify this, which bigram can distinguish with the accurate sentiment of the text.

The outcome of equation (6) considering all the words from text is shown in equation (7)

$$= \sum_{i=1}^n \prod_{j=1}^l \left[\begin{array}{c} (WD_{11}, WD_{13}, \dots, WD_{1l}) \\ (WD_{21}, WD_{23}, \dots, WD_{2l}) \\ \vdots \\ \vdots \\ \vdots \\ (WD_{n1}, WD_{n3}, \dots, WD_{nl}) \end{array} \right] \quad (7)$$

where $l = \frac{k}{2}$ as two words in the corpus are now combined into one word.

3.3 Distinguishing emotional features using profile of mood states (POMS)

There are two methods of handling semantic orientation recognition are manual analysis and lexicon based analysis. Manual analysis, though attains highest precision ingests costs, time and resources. Lexicon analysis, on the other hand, espouses the automated evaluation of sentiment analysis and thus commonly accepted (Wang et al., 2017). When constructing a lexicon, firstly, seed words are generated, which then identifies the mood state of responses grounded on these seed words (Mantyla et al., 2017). For the study, the lexicon analysis was done on the basis of POMS Scoring (Renger, 1992). The responses (in form of texts) are collected for six Hindi news channels and two English newspaper, amounts to 211,787 texts, are POMS-scored. The POMS-scoring function $P(t)$ maps each text to a six-dimensional mood vector $p \in R^6$ [*anger, depression, fatigue, vigour, tension, confusion*]. $P(t)$ matches the term extracted from of 211,787 texts to the set of POMS mood adjectives. POMS uses equation (7), and set the mood adjectives for dimensions i, j , as denoted by set p_{ij} . Thus, POMS scoring function can thus be defined as follows:

$$P(t) = p \in R^6 = \sum_{i=1}^n \prod_{j=1}^l \left[\begin{array}{c} (WD_{11} \cap p_{11}, WD_{13} \cap p_{13}, \dots, WD_{1l} \cap p_{1l}) \\ (WD_{21} \cap p_{21}, WD_{23} \cap p_{23}, \dots, WD_{2l} \cap p_{2l}) \\ \vdots \\ \vdots \\ \vdots \\ (WD_{n1} \cap p_{n1}, WD_{n3} \cap p_{n3}, \dots, WD_{nl} \cap p_{nl}) \end{array} \right] \quad (8)$$

The outcome of equation (8) can be then normalised to a vector p for each text t , to produce the unit mood vector,

$$\hat{p} = \frac{p}{\|p\|} \quad (9)$$

The benefit of transforming matrix into vector is that equation (9) can be effortlessly unravelled using Bayes theorem (considering A and B as two events for sample space) (Rrushi, 2011)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (10)$$

where

$P(A|B)$ is probability of A, when B is specified

$P(B|A)$ is probability of B, when A is specified

$P(A)$ is probability of A

$P(B)$ is probability of B.

$$= \frac{P(WD_1, WD_2, \dots, WD_l | T_1, T_2, \dots, T_l)}{P(WD_1, WD_2, \dots, WD_l)} \quad (11)$$

where $P(WD_1, WD_2, \dots, WD_l)$ is fixed and can thus be ignored.

$$= P(WD_1, WD_2, \dots, WD_l | T_1, T_2, \dots, T_l) \quad (12)$$

Applying the rule of product of probability, the final outcome is:

$$\prod_{i=1}^l P(WD_i | T_i) \cdot P(T_i | T_{i-1}) \quad (13)$$

In equation (13), $P(WD_i | T_i)$ is called emission probability and $P(T_i | T_{i-1})$ is called transition probability. The evaluation of equation (13) was accomplished using maximum likelihood estimate (MLE) method (Yilmaz and Pearson, 2015), depending on POMS analysis. $P(WD_i | T_i)$ represents the probability given a tag is associated with bigram and $P(T_i | T_{i-1})$ is computed by likelihood how often the first tag is substituted by the second tag. The representation is given in equation (14)

$$\prod_{i=1}^l \frac{C(T_i, WD_i)}{C(T_i)} \cdot \frac{P(T_{i-1}, T_i)}{C(T_{i-1})} \quad (14)$$

4 Research analysis

The research analysis includes the independent opinion analysis for all three movements, in order to establish the six feature set of POMS. The mood states feature comprises different textual opinion-related features for the time duration. Table 1 depicts the total number of text collected through 6,592 documents, with 211,787 texts generated after preprocessing.

Table 1 Total text collected after preprocessing

| <i>Total text</i> | <i>News published in YouTube</i> | | | | | <i>News published in news paper</i> | | | |
|-------------------|----------------------------------|-----------------|-------------|------------|-------------|-------------------------------------|---------------------------|---------------------------|--------------|
| | <i>Aaj Tak</i> | <i>ABP news</i> | <i>CNBC</i> | <i>IBN</i> | <i>NDTV</i> | <i>Zee news</i> | <i>The Indian Express</i> | <i>The Times of India</i> | <i>Total</i> |
| Demonetisation | 8,206 | 2,867 | 407 | 3,655 | 1,7561 | 13,089 | 28,225 | 11,662 | 85,672 |
| Digital India | 1,219 | 2,139 | 908 | 640 | 2,4393 | 1,753 | 17,460 | 5,337 | 53,849 |
| GST | 607 | 2,516 | 1,940 | 479 | 6,235 | 22,720 | 28,594 | 9,175 | 72,266 |

Table 2 Demonetisation results for six Hindi channels

| <i>Demonetisation</i> | <i>Ag</i> | <i>Dp</i> | <i>Ft</i> | <i>Vg</i> | <i>Ts</i> | <i>Cf</i> | <i>Total</i> |
|-----------------------|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| Aaj Tak | 36 | 312 | 24 | 9 | 8 | 59 | 448 |
| ABP news | 10 | 169 | 18 | 10 | 5 | 22 | 234 |
| CNBC | 1 | 26 | 0 | 2 | 1 | 5 | 35 |
| IBN | 13 | 99 | 4 | 1 | 3 | 15 | 135 |
| NDTV | 120 | 616 | 48 | 15 | 20 | 133 | 952 |
| Zee news | 78 | 593 | 62 | 35 | 23 | 112 | 903 |

Table 3 Digital India outcomes for six Hindi channels

| <i>Digital India</i> | <i>Ag</i> | <i>Dp</i> | <i>Ft</i> | <i>Vg</i> | <i>Ts</i> | <i>Cf</i> | <i>Total</i> |
|----------------------|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| Aaj Tak | 8 | 54 | 6 | 5 | 0 | 10 | 83 |
| ABP news | 16 | 117 | 6 | 5 | 3 | 14 | 161 |
| CNBC | 6 | 50 | 3 | 2 | 2 | 11 | 74 |
| IBN | 9 | 57 | 8 | 0 | 1 | 16 | 91 |
| NDTV | 59 | 301 | 22 | 13 | 10 | 55 | 460 |
| Zee news | 17 | 106 | 9 | 10 | 5 | 19 | 166 |

Table 4 GST results for six Hindi channels

| <i>GST</i> | <i>Ag</i> | <i>Dp</i> | <i>Ft</i> | <i>Vg</i> | <i>Ts</i> | <i>Cf</i> | <i>Total</i> |
|------------|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| Aaj Tak | 3 | 29 | 3 | 1 | 2 | 10 | 48 |
| ABP news | 13 | 114 | 16 | 10 | 4 | 28 | 185 |
| CNBC | 6 | 69 | 7 | 0 | 2 | 18 | 102 |
| IBN | 5 | 23 | 3 | 0 | 2 | 1 | 34 |
| NDTV | 43 | 224 | 29 | 6 | 7 | 44 | 353 |
| Zee news | 123 | 1,248 | 98 | 60 | 64 | 196 | 1,789 |

Table 5 Demonetisation results for two English news papers

| <i>Demonetisation</i> | <i>Ag</i> | <i>Dp</i> | <i>Ft</i> | <i>Vg</i> | <i>Ts</i> | <i>Cf</i> | <i>Total</i> |
|-----------------------|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| Times of India | 6 | 29 | 4 | 1 | 0 | 5 | 45 |
| The Indian Express | 12 | 64 | 1 | 2 | 3 | 10 | 92 |

Table 6 Digital India results for two English news papers

| <i>Digital India</i> | <i>Ag</i> | <i>Dp</i> | <i>Ft</i> | <i>Vg</i> | <i>Ts</i> | <i>Cf</i> | <i>Total</i> |
|----------------------|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| Times of India | 0 | 18 | 4 | 2 | 2 | 2 | 28 |
| The Indian Express | 4 | 23 | 7 | 2 | 1 | 7 | 44 |

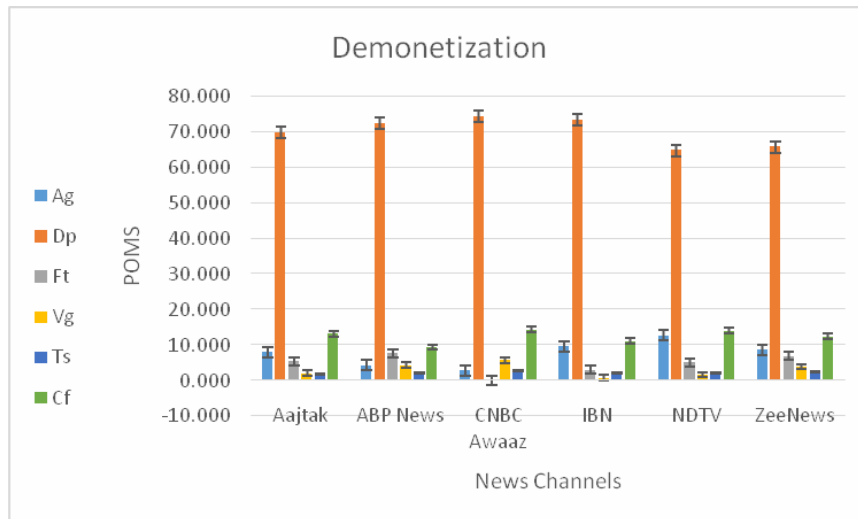
Table 7 GST outcomes for two English news papers

| <i>GST</i> | <i>Ag</i> | <i>Dp</i> | <i>Ft</i> | <i>Vg</i> | <i>Ts</i> | <i>Cf</i> | <i>Total</i> |
|--------------------|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| Times of India | 8 | 22 | 4 | 1 | 2 | 1 | 38 |
| The Indian Express | 13 | 51 | 13 | 2 | 4 | 9 | 92 |

The study endorses the conclusion that the citizens’ mood state is a kind of psychosomatic characteristic that could be inevitably learned through their social network behaviour. The result also indicates that news published by news channels and news papers online, plays an important role in calculating users’ mood state.

Tables 2–7 display the mood states of citizen, including the total number of documents for all three movements.

Figure 2 Demonetisation and response analysis for six Hindi news channel (see online version for colours)



The distinguishable factor for all eight media is *depression*. The outcomes, as mentioned in Tables 2–7 evidently specifies $\{depression\} \ll \{anger \cup fatigue \cup vigour \cup tension \cup confusion\}$. Public opinion based on all these eight media was exceedingly deleterious

about all the three movements and there was no evidence available which directed of any communication strategies framed by GOI to tackle this dint of news.

Based on Table 1–6, Figures 2–7 depicts the outcomes in terms of percent of responses for all three movements in all eight online media.

Figure 3 Digital India and response analysis for six Hindi news channel (see online version for colours)

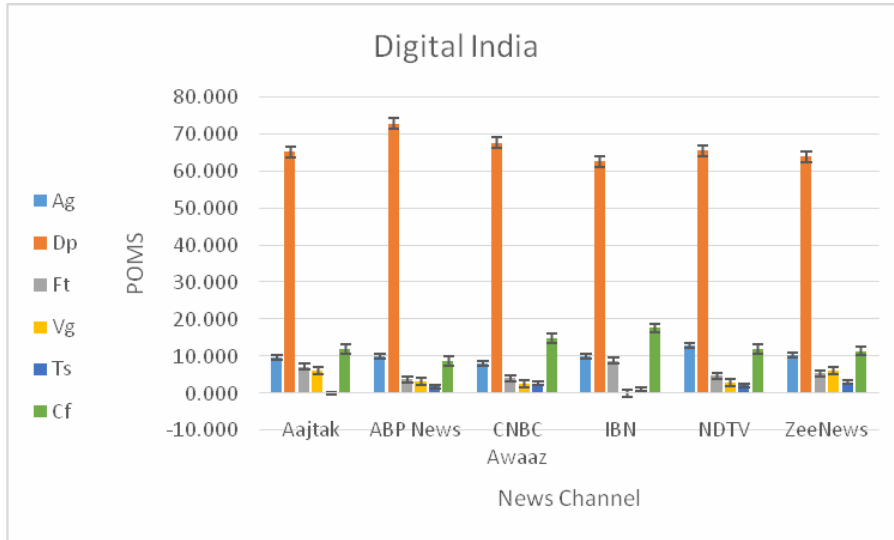


Figure 4 GST and response analysis for six Hindi news channel (see online version for colours)

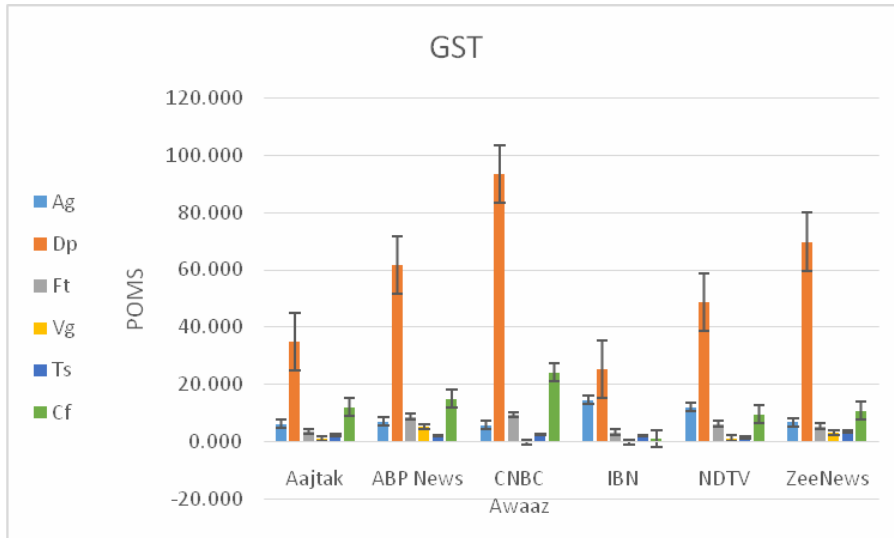


Figure 5 Demonetisation and response analysis for two English news paper (see online version for colours)

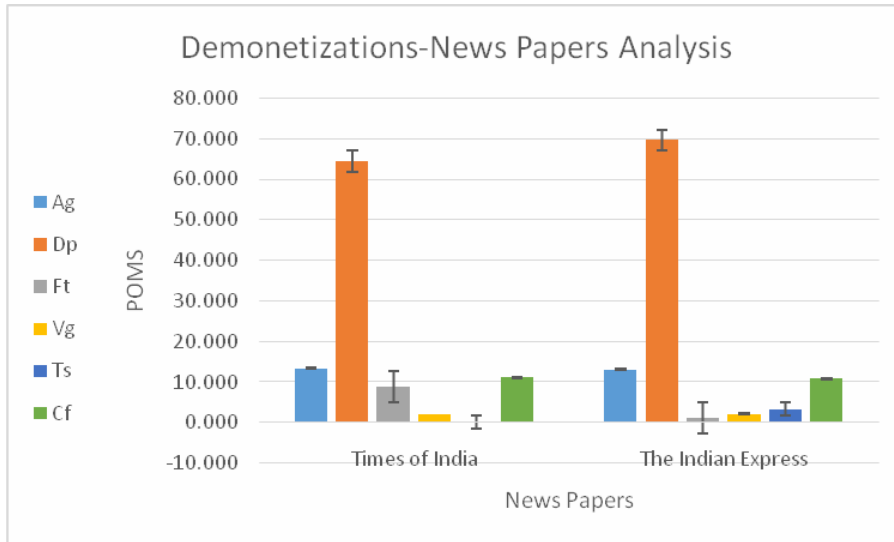


Figure 6 Digital India and response analysis for two English news papers (see online version for colours)

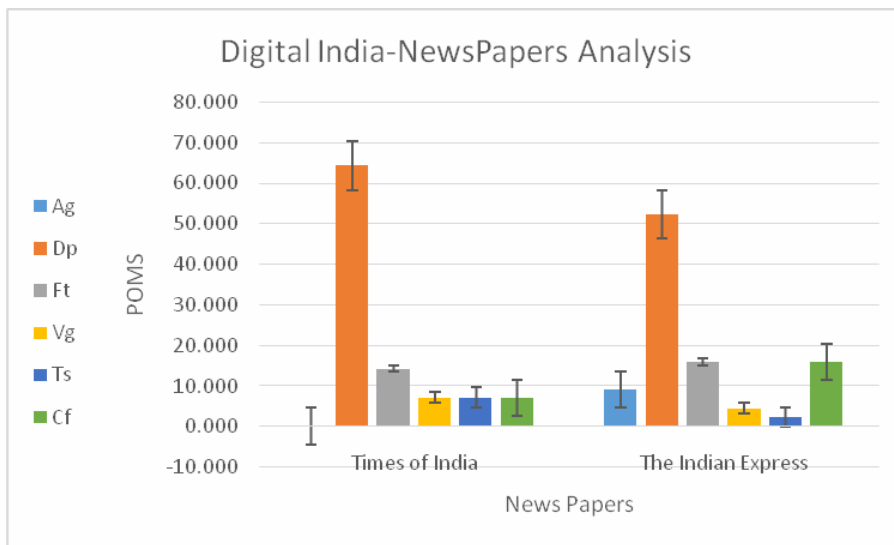
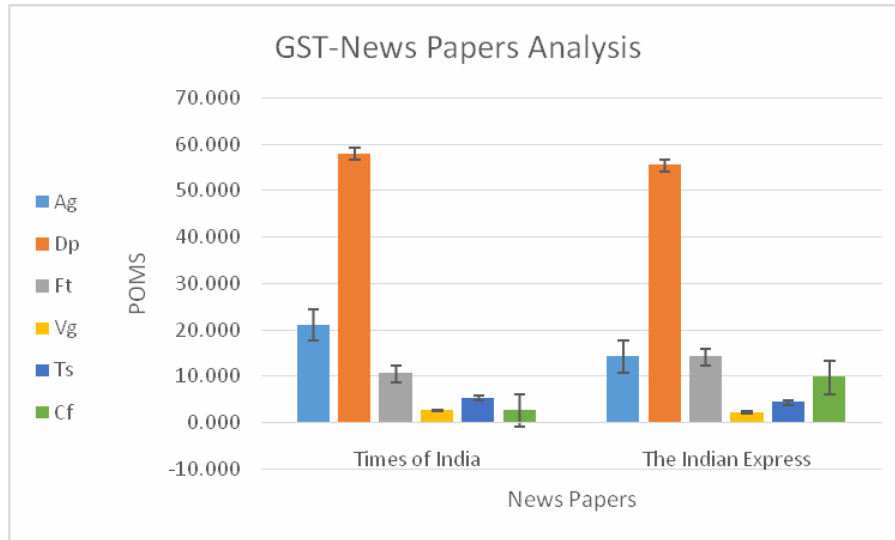


Figure 7 GST and response analysis for two English news papers (see online version for colours)

5 Inferences and conclusions

This study proposed POMS analysis method to calculate mood states of online users repeatedly, assuming natural language processing and the machine learning method. Some assumptions can be drawn through analysis and discussion. Through the analysis of users' (public) POMS scores and their opinions articulated in news published online, the study found that the user's mood state is correlated to the text expression. Users' emotional expressions on the online medium played an imperative part in the process of obtaining their factual mood states. In-depth analysis and vigilant observation also exhibited that opinion remain persistent even on diverse time span with users' mood states. However, there were still some restrictions in the study, as only eight online media were selected and characteristics of mood states vis-à-vis those with diverse identities and ages were not deliberated. In addition, this paper constructed a mood state lexicon to extract the citizens' opinion analysis using POMS tagging. The scale is reasonably small and the feature extraction process for the mood states of users under large-scale data necessitates further research.

References

- Al-Ayyoub, M., Essa, S.B. and Alsmadi, I. (2015) 'Lexicon-based sentiment analysis of Arabic tweets', *International Journal of Social Network Mining*, Vol. 2, No. 2, pp.101–1114, <https://doi.org/10.1504/IJSNM.2015.072280>.
- AleEbrahim, N. and Fathian, M. (2013) 'Summarising customer online reviews using a new text mining approach', *International Journal of Business Information Systems*, Vol. 13, No. 3, pp.343–358, <https://doi.org/10.1504/IJBIS.2013.054468>.
- Antilla, L. (2014) 'Climate change in the media: reporting risk and uncertainty', *Environmental Communication*, Vol. 8, No. 4, pp.551–553, <https://doi.org/10.1080/17524032.2014.980107>.

- Bollen, J., Pepe, A. and Mao, H. (2009) 'Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena', Paper presented at the *Meeting of WWW 2009 Conference*, Madrid, Spain.
- Booth, N.M. (2011) 'Mapping and leveraging influencers in social media to shape corporate brand perceptions', *Corporate Communication: An International Journal*, Vol. 16, pp.184–191.
- Eryürük, S., Koshijima, I. and Kato, T. (2014) 'Impact analysis of political decisions on business sustainability', *Asian J. of Management Science and Applications*, Vol. 1, No. 4, p.355, <https://doi.org/10.1504/AJMSA.2014.070664>
- Füller, J. (2010) 'Refining virtual co-creation from a consumer perspective', *California Management Review*, Vol. 52, No. 2, pp.98–122, <https://doi.org/10.1525/cmr.2010.52.2.98>.
- Ganesan, K., Zhai, C. and Viegas, E. (2012) 'Micropinion generation: an unsupervised approach to generating ultra-concise summaries of opinions', in *Proceedings of the 21st International Conference on World Wide Web*, ACM, New York, NY, USA, pp.869–878 [online] <https://doi.org/10.1145/2187836.2187954>.
- Haddi, E., Liu, X. and Shi, Y. (2013) 'The role of text pre-processing in sentiment analysis', *Procedia Computer Science*, Vol. 17, pp.26–32, <https://doi.org/10.1016/j.procs.2013.05.005>.
- HT (2015) 'Govt launches Digital India initiative with Twitter Samvad', *Hindustan Times*, 25 March [online] <https://www.hindustantimes.com>.
- Kermanidis, K.L. and Maragoudakis, M. (2013) 'Political sentiment analysis of tweets before and after the Greek elections of May 2012', *International Journal of Social Network Mining*, Vol. 1, Nos. 3–4, pp.298–317, <https://doi.org/10.1504/IJSNM.2013.059090>.
- Klotz, C., Akinalp, C. and Unger, H. (2016) 'Clustering user behaviour patterns on Twitter', *International Journal of Social Network Mining*, Vol. 2, No. 3, p.203 <https://doi.org/10.1504/IJSNM.2016.082632>.
- Kumar, N., Srinathan, K. and Varma, V. (2016) 'A graph-based unsupervised N-gram filtration technique for automatic keyphrase extraction', *International Journal of Data Mining, Modelling and Management*, Vol. 8, No. 2, pp.124–143, <https://doi.org/10.1504/IJDM.2016.077198>.
- Liu, B., Hu, M. and Cheng, J. (2005) 'Opinion observer: analyzing and comparing opinions on the web', in *Proceedings of the 14th International Conference on World Wide Web*, New York, NY, USA, ACM, pp.342–351, <https://doi.org/10.1145/1060745.1060797>.
- Malley, B., Ramazzotti, D. and Wu, J.T. (2016) 'Data pre-processing', in *Secondary Analysis of Electronic Health Records*, pp.115–141, Springer, Cham, https://doi.org/10.1007/978-3-319-43742-2_12.
- Manning, C.D., Raghavan, P. and Schütze, H. (2008) *Introduction to Information Retrieval*, Cambridge University Press, New York.
- Mantyla, M.V., Novielli, N., Lanubile, F., Claes, M. and Kuutila, M. (2017) 'Bootstrapping a lexicon for emotional arousal in software engineering', in *Proceedings of the 14th International Conference on Mining Software Repositories (MSR '17)*, IEEE Press, Piscataway, NJ, USA, pp.198–202, <https://doi.org/10.1109/MSR.2017.47>.
- Menon, B. (2017) *YouTube Mobile Viewership Hits 180 m in India – Business Line*, 03/23 [online] <https://www.thehindubusinessline.com/info-tech/youtube-mobile-viewership-hits-180-m-in-india/article9598547.ece> (accessed 10 May 2018).
- Naaman, M., Boase, J. and Lai, C-H. (2010) 'Is it really about me?: Message content in social awareness streams', in *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, ACM, New York, NY, USA, pp.189–192 [online] <https://doi.org/10.1145/1718918.1718953>.
- Porter, M.F. (1997) *Readings in Information Retrieval*, in Sparck Jones, K. and Willett, P. (Eds.): pp. 313–316, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA [online] <http://dl.acm.org/citation.cfm?id=275537.275705> (accessed 2 May 2018).

- Renger, R. (1992) 'A review of the profile of mood states (POMS) in the prediction of athletic success', *Journal of Applied Sport Psychology*, Vol. 5, No. 1, <https://www.tandfonline.com/doi/abs/10.1080/10413209308411306> (accessed 4 May 2018).
- Rrushi, J.L. (2011) 'A Bayesian theory of confirmation for intrusion report fusion in process control networks', *International Journal of Critical Computer-Based Systems*, Vol. 2, No. 2, p.162 <https://doi.org/10.1504/IJCCBS.2011.041258>.
- Shahid, A., Wilkinson, K., Marcu, S. and Shapiro, C.M. (2011) 'Profile of mood states (POMS)', in *STOP, THAT and One Hundred Other Sleep Scales*, pp.285–286, Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9893-4_68.
- Sinickas, A. (2011) 'Who needs surveys when you scan social media', *Strategic Communication Measurement*, December 2010/January 2011, Melcrum Publishing Ltd., London, UK [online] Retrieved from <https://www.sinicom.com/resources/publications/surveys-vs-social-media/> (accessed 29 April 2018).
- Sreeja (2016) *Emotion Recognition from Poems by Maximum Posterior Probability* [online] https://www.researchgate.net/publication/309463787_Emotion_Recognition_from_Poems_by_Maximum_Posterior_Probability (accessed 11 May 2018).
- Subitha, S. and Sujatha, S. (2017) 'Context-based information retrieval from large heterogeneous data sources using semantics and polarity-based ranking', *International Journal of Computer Aided Engineering and Technology*, Vol. 9, No. 4, pp.497–507, <https://doi.org/10.1504/IJCAET.2017.086957>.
- Tan, C-M., Wang, Y-F. and Lee, C-D. (2002) 'The use of bigrams to enhance text categorization', in *Inf. Process. Manage*, Vol. 38, No. 4, pp.529–546.
- Wang, W., Li, Y., Huang, Y., Liu, H. and Zhang, T. (2017) 'A method for identifying the mood states of social network users based on cyber psychometrics', *Future Internet*, Vol. 9, No. 2, p.22, <https://doi.org/10.3390/fi9020022>.
- Yan, Y. (2011) (2) *Social media in Branding: Fulfilling a Need* [online] https://www.researchgate.net/publication/233643484_Social_media_in_branding_Fulfilling_a_need (accessed 5 May 2018).
- Yilmaz, H. and Pearson, N.D. (2015) 'Maximum likelihood estimation of covariance matrices with constraints on the efficient frontier', *International Journal of Computational Economics and Econometrics*, Vol. 6, No. 1, pp.71–92, <https://doi.org/10.1504/IJCEE.2016.073352>.