DEVELOPING A BI-LINGUAL LEXICON TO MAP CONSUMER PERCEPTION OF INDIAN BRANDS

Aditi Mudgal, Birla Institute of Management Technology, India[1]
Amrendra Pandey, Birla Institute of Management Technology, India
Amarnath Bose, Birla Institute of Management Technology, India
Pankaj Priya, Birla Institute of Management Technology, India

EXTENDED ABSTRACT

The way brands connect with consumers in the new era of digitalisation has redefined the traditional ways of marketing. In the current times, a consumer relies heavily on digital interactions (Edelman, 2010). Social media facilitates the consumer to consult with buyer instantly (Christodoulides & Jevons, 2011) and express their perceptions with each other. A report by Statista on Online video and streaming (2017) suggests that users are increasingly engaging themselves in snackable videos while on the move every day. Hence, online narrative videos are adopted by Indian brands as their key marketing tools. The emotions that occur during viewing the narrative videos should be identified and measured. This will help in understanding the perception formed for the brands advertised. The objective of the study, is to develop a bilingual lexicon for Indian social media landscape for mapping perception formed by the online video narrative advertisements.

Setting the context (Premise)

With the introduction of social media, new ways of interaction have emerged between brand and consumers (Hansen, Shneiderman, & A, 2011). According to the EY Report (2016), social media brings the key challenges for marketers such as a) sustaining or increasing engagement rates, b) content creation and c) measuring effectiveness. For Indian brands the scenario at digital ecosystem is relatively new (Senthil, Prabhu, & Bhuvaneswari, 2013). They are becoming approachable to customers through interactive Facebook pages and YouTube channels (KPMG India, 2017). User generated content forces smart marketers to listen. Consumer perception is the key indicator that establishes link between the consumer's interest and purchase behaviour of the consumer (Ali, Shabbir, Rauf, & Hussain, 2016). If we talk about, India as a context of study, we as researchers should essentially remember the degree of cultural diversity it has, so are its languages and vernaculars. The language used by 'Indian netizens' in digital landscape is not always in pure English. This is concoction of two languages is popularly known "Hinglish" which is defined as the mix of two primary languages, Hindi and English (Kothari & Snell, 2011) respectively. The unavailability

---

[1] aditi.mudgal_fpm16@bimtech.ac.in

of any tool present to analyse and standardise these bundle of opinions written on social media in hinglish is a huge gap and hinderance to fruitful insights for Indian marketing fraternity. Loosing out on Hinglish words used by users, erodes the essence of the comment. While analysing many times, the whole comment is categorised as null and assigned zero as a score. Hence as researchers we intend to fill in this gap by constructing a bi-lingual lexicon for analysing texts which are written in hinglish verbatim or consists of colloquialism or any other cultural referent.

Lexicons

With the accessibility of web and social media, people are independent to bring their opinions, grievances on various platforms. Hence, marketing research these days is extremely dependent on platforms such as Twitter and Facebook to capture consumer sentiment about brands and products. Analysing these opinions towards brand will certainly help the brand to chalk out their future marketing strategies (Hu & Liu, 2004). The complex factor pertaining to analysis of informal web texts, classified as 'sentiment' detection has been dealt with several techniques, which resulted in General and Subjective lexicons. Lexicons are standardised vocabularies that facilitate communication across diverse audiences. The Lexicons are scientific vocabularies which expand the standardisation of sensory science. The need for lexicons arose from the complex needs of industry (Lawless & Civille, 2013). Table 2 represents the previous works done on lexicon development in different areas of management

Table 2: Existing English resources for sentiment polarity and emotion

The essential traits of lexicons are consistency and precision in identifying sentiments. Emotional evaluation of a word at semantic level is called a sentiment. Emotion attributes have been generally associated with brands but little work has been published to understand consumer emotions associated with the brand itself. Lexicon use the mechanism of emotion profiling to analyse sentiments of reviews, blogs etc posted on social media. Through this research we try to construct a lexicon specifically for marketing domain using online narrative ads, to capture the consumer perception towards brands. We will be using bootstrapping technique to annotate words to different semantic and emotion subsets, a dataset of user comments on narrative ads were mined from Facebook official pages of selected brands, which will further analysed for sentiment detection and scoring according to the varying degrees of emotions represented.

Data Collection

Platform: Social media Platform - Facebook
Type of Data: Secondary Data
Unit of Analysis: Comments on Narrative videos of selected brands on the brands official Facebook page.

Themes/Kinds of Advertisement:
Online narrative advertisements which uses the story to build experiences around it brand's values and ethos. Narrative advertising appeals the consumer"s affective and emotional responses. These online narrative should not exclusively talk about the brand they endorse, this being the key characteristic of the narratives selected. The reasoning behind the selection of such narratives is an attempt to remove any prior biases towards the brand advertised, so that the consumer's opinion only talk about the brand and the content of narrative in particular.

Methodology

Step 1 Campaign Identification
Identification of narrative campaigns published on social media by selected Indian brands
Step 2 Comment Extraction (Unit of Analysis - Comment; Category -> Brand -> Campaigns)
Through the use of R open-source software, using Facebook API
Step 3 Data Cleaning (Stemming)
Step 4 Lexicon Building
Step 4 (i) Extraction of n-gram, where n =1,2,3 | Create a list of words for each Campaign
Step 4 (ii) Annotation. (Industry expert, students, commoners to get classification of each n grams)
Step 5 Using the Lexicon, assign sentiment score Category | Brand | Campaign wise
Step 6 Analyse sentiment trends differences across Category and Time

As researcher we intend to develop lexicon based on narrative advertisements for Indian brands, through the extracted comments of the respective online video narratives. The study aims at inferring the brand perception of consumers, this insight will add to the breadth of knowledge on brands and help refine future online marketing strategies in accordance to audiences tastes. We will be using Hu & Liu lexicon to conduct the analysis because this lexicon has been used successfully in similar projects. Hu & Liu's lexicon has obtained an average accuracy for the five products is 84%. Which is remarkable for any lexical analysis tool. Also this study intends to consider n-grams which is a phrase, or a combination of words which make sense.

The objective of using n-grams is to keep the essence intact while analysing and categorising these words in positive and negative. If we go by the Oxford Advanced Learner's Dictionary, Hinglish is a language which is a mixture of English and Hindi, especially a type of English that includes many Hindi words. Initially bootstrapping technique would be employed to categorise the words into positive and negative sentiments, the reasoning behind using bootstrapping is that categorisation of words on the basis of critical understanding by Indian consumers and population is useful in

introducing new words into a lexicon. The annotators can be student who represents a majority of Indian online population, while the market and industry experts can be helpful in understanding the technical representation of word according to consumer behaviour domain. Sentiment profiling of words is done by categorising them into positive and negative using bootstrapping.

Analysis and Results

We are working and beginning our analysis process. We are currently on step 4(ii) mentioned above. Annotation is a very lengthy and humungous task. This task determines the strength and accuracy of lexicon, hence it is one of the core responsibility of the researchers while building a lexicon, as it lays the foundation of the outcome. Our corpus data deals with 80,000 total comments.

Which comprises of 8161 english words and 35,056 non-english words. The number of non-english words is four times the english words which shows the large gap of un-analysed data. This also represents the pressing need of building a bi-lingual lexicon for Indian context.

Lexicon such as SentiHealth.general purpose lexicon, consists of words related to health and hence limits itself to specific domain, it is based on hybrid approach for efficient classification. Sentiment classification can usually be seen as a two-class classification problem. In this type of research, sentiment analysis is essentially a text classification problem, which involves semantics.

In current academic literature sentiment analysis studies have either concentrated on enumerating technical details in the areas of social media analysis, text mining, natural language processing, and data mining or just focus on a certain aspects of sentiment analysis. Our process of developing a tool to analyse bi-lingual sentiments derives the characteristics from majority of studies, primarily focusing on key characteristics of Hu & liu's lexicon, to which, in a manner we are extending and adding too.

We would be presenting our research from new perspectives of task-orientation, granularity by studying ngrams and uni-grams and methodology. We will state how our lexicon will help in analysing sentiment with much more finer approach. Through out the journey of developing a bilingual lexicon we intend to walk the reader through the motivations, benefits, and limitations of the new lexicon and explore the technicalities associated with techniques and methods used in preparing it.

While the study is still being conducted, we have strong results on the granularity and sharper results through different lexicon analyis. Currently we are working on the results derived for themes of video narrative and the subsequent perception formed by these advertisements.

Conclusions

The objective was to construct a bi-lingual lexicon with their respective polarity and degrees of the referred sentiment of word. This bi-lingual lexicon is structured out of a large number of consumer opinions of the various brands advertised online. We believe that this problem will become increasingly important as more people are consuming online narratives while building perception through these narratives and simultaneously expressing their opinions on the web. We believe that this lexicon may be used in practical settings of social media dialogue exchanges to bring better insights in the area online consumer psychology broadly. (consumer behaviour and interactivity exercises)

Acknowledgements

References

Ali, Z., Shabbir, M. A., Rauf, M., & Hussain, A. (2016). To Assess the Impact of Social Media Marketing on Consumer Perception. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, *6*(3), 69-77

Christodoulides, G., Jevons, C., & Blackshaw, P. (2011). The voice of the consumer speaks forcefully in brand identity: User-generated content forces smart marketers to listen. *Journal of Advertising Research*, *51*(1 50th Anniversary Supplement), 101-111

Edelman, D. C. (2010). Branding in the digital age. *Harvard business review*, *88*(12), 62-69

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345-360

Social Media Marketing: India Trends Study 2016 - EY Report

Hu, M., & Liu, B. (2004, July). Mining opinion features in customer reviews. In *AAAI* (Vol. 4, No. 4, pp. 755-760)

Lawless, L. J., & Civille, G. V. (2013). Developing lexicons: A review. *Journal of Sensory Studies*, *28*(4), 270-281

Kiritchenko, S., Zhu, X., Cherry, C., & Mohammad, S. (2014). NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (pp. 437-442)

KPMG India - FICCI Media and Entertainment Report 2017.

Kothari, R., & Snell, R. (Eds.). (2011). *Chutnefying English: The Phenomenon of Hinglish*. Penguin Books India

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2011). SentiFul: A lexicon for sentiment analysis. *IEEE Transactions on Affective Computing*, *2*(1), 22-36

Rodrigues, E. M., Milic-Frayling, N., Smith, M., Shneiderman, B., & Hansen, D. (2011, October). Group-in-a-Box layout for multi-faceted analysis of communities. In

Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on (pp. 354-361). IEEE.

Senthil, M., Prabhu, N. R. V., & Bhuvaneswari, S. (2013). Customers' perception towards advertising in the online shopping and social networking websites among Internet users in India'. *AMET, Internafional Journal of Management*, 2(1), 50-59.

Strapparava, C., & Valitutti, A. (2004, May). Wordnet affect: an affective extension of wordnet. In *Lrec* (Vol. 4, pp. 1083-1086).