Sentimental Analysis For Electronic Product Review

Kaustub Shetty, Shravan Thombre, Ranjeet Yadav, Dheeraj Prajapati

Kaustub Shetty, Atharva college of engineering, Maharashtra, India
Ranjeet Yadav, Atharva college of engineering, Maharashtra, India
Shravan Thombre, Atharva college of engineering, Maharashtra, India
Dheeraj Prajapati, Atharva college of engineering, Maharashtra, India

---

Abstract - *Due to its numerous uses and applications, sentiment analysis has become one of the most well-known study areas. Business analytics places a high priority on monitoring social media, forums, blogs, and other online resources for consumer evaluations, product competition, and survey replies to discover customer insight. The usage of mixed language has become widespread due to the increase in informal user provided data online. Linguistic code switching, or the habit of utilizing more than one language in a single sentence, is one way that mixed language develops. Prior to now, sentiment analysis has rarely been used in language that is so mixed-up. The prior methods of sentiment analysis for such material are useless due to the unclear grammatical structure. In this study, we suggest a Describe method for analyzing the tone of sentences written in a language that combines Hindi and English lexicons. Our method may be applied to mixed language data as well as data from any one of the source languages to assess sentiment. Sentiment analysis has taken into consideration the many grammatical transitions found in mixed-language texts. We provide case studies with social media data sets to show the usefulness of the suggested strategy.*

Key Words: *Sentiment, Codemix*

1. INTRODUCTION

The sentiment analysis is the process of extracting irrational information from a source using computational linguistics and natural language processing methods [1]. Research in sentiment analysis is motivated by the rise of user-generated material online. Opinion mining has greatly gained interest in the scientific community [2] as a result of the growth of online resources with a high concentration of opinions, such as blogs, social media, and review websites [1]. businesses seeking to advertise themselves. Instead of spending money and effort on long-consuming and expensive market research projects, companies may use internet opinion mining to manage their reputation [3]. Many studies have been conducted to comprehend mixed language since individuals frequently use it in social media to convey their viewpoints

A mixed language is one that results from the mixing of typically two or more source languages, making it hard to categorize the new language as belonging to any of the original language groups [8]. This method is often referred to as Linguistic Code Switching (LCS) [5]. There is a dearth of research being done on sentiment analysis of material that is written in many languages, despite the fact that sentiment analysis of texts that are solely written in one language has received considerable attention [1]. The sentiment analysis for mixed language is a new area of research for us.

In this work, Hindi and English are combined to create a hybrid language, and a method for sentiment analysis of that language is discussed. The resultant grammar in the hybrid language typically switches back and forth between the parent languages and departs greatly from their grammatical structures. In order to evolve the overall sentiment of a phrase in the mixed language, our technique takes into account the determination of the grammatical transition and the application of sentiment combination rules across languages. Results from experiments validate the efficacy of the suggested approach.

2. RELATED WORK

A four step approach involving language identification, part-of-speech tagging, subjectivity detection and polarity detection has been deployed in opinion mining of multilingual data in social media [9]. This method may be ineffective for data which uses linguistic code switching. In LCS a sentence contains multiple languages in random order and hence only word by word language identification can be done. This however may lead to loss of context which is crucial in sentiment analysis [10].

Multilingual sentiment analysis which requires training a classifier on different languages is explored. A Naïve Bayes classifier has been trained on source sets of millions of tweets in English, German, French and Portuguese to detect the sentiment polarity. It involves a semi-supervised heuristic labeling scheme to acquire large amounts of training data in a variety of languages, and content-based features that work well across languages [13].

Another work in the domain of multilingual sentiment analysis uses stylistic and syntactic features of Arabic and English language for classification and opinion mining of English and Arabic web forms. Specific feature extraction components have been integrated to account for linguistic characteristics of Arabic [14].

3. PROPOSED WORK

This section presents our proposed code-mixed sentimental analysis framework. Product reviews contents from popular Indian e-commerce sites like amazon.in, flipkart.com, snapdeal.com, shopclues.com are collected as our dataset. The dataset has a lot of typos in the form of joint words like "very good" as well as abbreviations containing numerals such as "gr8" for great. One of the primary focuses of this work is to pre-process the product review available on Indian E-commerce sites so that the reviews contain only English text. Once reviews are converted to English text, the reviews are stored in the database. These reviews are presented with detailed analysis to the admin through help of these reviews admin can analyze and can make important business decision regarding improving the bad reviewed product. These will help the seller to automate manual task which is time consuming.

The main contributions of the proposed research are the following:

- The generation of a sentiment score using a lexicon-based approach for each product review of the dataset.
- Labeling the review texts as negative if the generated sentiment score is <0 or positive if the score is >1.
- The combination of all product reviews into a single data frame to obtain more sentiment-related words.
- Improving the accuracy by developing a l LSTM model for the product-related sentiment classification.

4. DATA PREPROCESSING

The data preprocessing is done using NLP(Natural Language processing) which includes following steps

A. Removing Urls

The user generated comments may contain some bad links or the links which are not related to our domain. So it is useful to remove those links for preprocessing optimization

B. Text Translation

The comments may contain code-mix data which is the data in the mix language format i.e. in Hindi and English. So it is necessary to convert the code-mix language to one specific language (English) in order to perform further steps.

C. To Lowercase

To convert user comments to one specific format i.e. to lowercase

D. To Remove Extra White Spaces

User comments might contain some extra blank lines or many spaces which are of no use. So it is essential to remove those whitespaces which will save some memory

E. To remove Punctuation

Remove punctuations as they are not contributing to predicting overall product sentiments

F. Spell Correction

User comments may contain misspelled words which might result in some different result during prediction. So it is necessary to correct those misspelled words

5. PROPOSED MODEL

LSTM is a type of RNN capable of learning long-term dependence [52]. We used an LSTM layer and assigned it to 50 hidden units toward the next layer. One of the most notable advantages of employing a convolutional neural network as feature extraction technique beyond a traditional LSTM is the reduction in the aggregating amount of features. Throughout the feature extraction process, a sentiment classification model uses these features (words) for prediction of the product review text as positive or negative sentiment. LSTM executes pre calculations for the input sequences before providing an output to the last layer of the network. In every cell, four discrete computations are conducted based on four gates: input(it), forget (ft), candidate(ct), and output(ot). The structure of the LSTM model is presented in Figure 3. The equations for these gates are as follows:

$$\begin{aligned}
f_t &= \operatorname{sig}\left(Wf_{xt} + Uf_{ht} - 1 + b_f\right),\\
i_t &= \operatorname{sig}\left(Wi_{xt} + Ui_{ht} - 1 + b_i\right),\\
O_t &= \operatorname{sig}\left(Wo_{xt} + Uo_{ht} - 1 + b_o\right),\\
c \sim t &= \tanh\left(wc_{xt} + Uc_{ht} - 1 + bc\right),\\
C_t &= \left(f_{to}ct - 1 + i_{to}c \sim t\right),\\
h_t &= O_{to} * \ \tanh\left(C_t\right),\\
\tanh(x) &= \frac{1 - e^{2x}}{1 - e^{2x}},
\end{aligned} \tag{1}$$

where sig and tanh are the sigmoid and tangent activation functions, respectively, X is the input data, W and b represent the weight and bias factor, respectively, Ct is the cell state, c ∼ t is the candidate gate, and refers to the output of the LSTM cell.

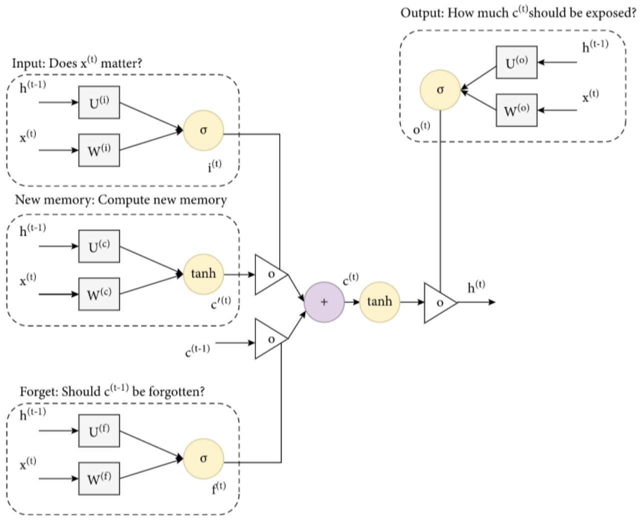

Fig 1: Structure of LSTM

6. PROPOSED SYSTEM ARCHITECTURE

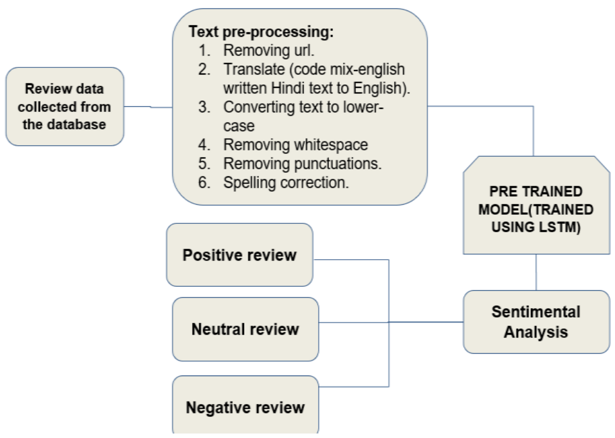

Fig 2: Proposed system architecture

- The reviews written by Indian buyers are mainly in English, but it contains some Hindi texts (written in English Scripts only).

- The proposed system performs some text preprocessing on the provided review. It includes following text processing:

  - Removing url.
  - Translate (code mix-english written Hindi text to English).
  - Converting text to lower-case
  - Removing whitespace
  - Removing punctuations.
  - Spelling correction.

- The preprocessing helps reduce the effort for sentimental analysis by eliminating stopwords and extracting keywords required for sentiment analysis. The text is then passed to the pretrained model (trained using LSTM) which will provide the result accordingly

7. RESULTS

We trained three models for performing sentimental analysis namely SVM(Support Vector Machine), LR (Logistic Regression), LSTM(Long Term Short Memory) and used same on all three models to do comparative analysis

This Dataset is a labeled dataset which consists of 938,254 customer reviews which are rated from 1 to 5 star rating, Further we segregated these reviews into three categories positive, negative and neutral.

Following are the results achieved by the comparative analysis of these three models.

Model	Parameters	Precision	Recall
SVM	Positive	78.55	76.80
	Neutral	65.03	58.56
	Negative	77.67	83.67
LOGISTIC REGRESSION	Positive	78.22	80.92
	Neutral	63.44	59.43
	Negative	82.53	82.46
LSTM	**Positive**	**83.64**	**81.38**
	Neutral	**87.47**	**82.66**
	Negative	**88.09**	**89.03**

Table -1: comparative analysis of different model

Among the above tested models, the LSTM provided the highest overall accuracy due to which we used the LSTM model for our sentimental analysis.

Following are expected and actual results of sentimental analysis provided by the LSTM model for a given customer review provided that the text given by the customers are preprocessed before giving them as input to the model for further sentimental analysis

Text	Expected	Predicted
Received the product in good condition. Writing review after one hour of usage. Picture quality is perfect and sound is too good. Connectivity is fast enough \\][. Overall worth for the price. Sound is perfect for medium sized hall and we should have HD pack for best viewing experience	Positive	Positive
Camera thik thak hai ...!!! HAD A VERY AVERAGE EXPERIENCE KOI AUR PHONE DEKH LO BHAI..	Neutral	Neutral
प्रोडक्ट बुहत गन्दा है	Negative	Negative
Overall design acchi nahi but camera is worth buying	Negative	Neutral

Table -2: actual and expected output of LSTM MODEL

8. CONCLUSIONS

The proposed method focuses on sentiment analysis both at the phrase level and at the sub-phrase level of a sentence. This is important in mixed-language sentences that consist of syntactic combinations from the source language. The classifier is trained on mixed language training data to learn grammatical transitions that often occur in mixed languages. Finally, sentiment combination rules are used to evaluate the sentiment of phrases in different languages and predict the overall sentiment of mixed-language sentences.

In this method, the sentimental classifier learns simplified grammatical structures of mixed languages. This is sufficient for parsing short sentences like those found on social networking sites and tweets. This provides an alternative to the trivial method of applying sentiment analysis after translating mixed-language text written from one language to another. Future work will expand the scope of the experiment to include text samples in more than one language. We also plan to experiment with larger training and test sets, as well as more advanced methods such as uniform grouping of phases and subphases to further improve accuracy.

9. REFERENCES

[1] G. Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An Enhanced lexical resource for sentiment analysis and opinion mining. In LREC,volume 10, pages 2200–2204, 2010

[2] Agarap AF, Grafilon P (2018) Statistical analysis on E-commerce reviews, with sentiment classification using bidirectional recurrent neural network (RNN). arXiv preprint arXiv:1805.03687

[3] Ain QT, Ali M, Riaz A, Noureen A, Kamran M, Hayat B, Rehman A (2017) Sentiment analysis using deep learning techniques: a review. Int J Adv Comput Sci Appl 8(6):424

[4] Daassi-Gnaba H, Oussar Y (2015) External vs. internal svm-rfe: the svm-rfe method revisited and applied to emotion recognition. Neural Network World 25(1):75

[5] Fang X, Zhan J (2015) Sentiment analysis using product review data. Journal of Big Data 2(1):5

[6] Gupta, S. (2018, August 06) Applications of sentiment analysis in business.

[7] He R, McAuley J (2016, April) Ups and downs: modeling the visual evolution of fashion trends with oneclass collaborative filtering. In proceedings of the 25th international conference on world wide web (pp.507-517). International world wide web conferences steering committee

[8] Hochreiter S (1998) The vanishing gradient problem during learning recurrent neural nets and problem solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 6(02):107–116

[9] Huang JP, Wang XA, Zhao Y, Xin C, Xiang H (2018) Large earthquake magnitude prediction in Taiwan based on deep learning neural network. Neural Network World 28(2):149–160

[10] Ilayaraja S (2018, June 26) ilayaraja97/keras-contrib: Keras community contributions

[11] Jadav BM, Vaghela VB (2016) Sentiment analysis using support vector machine based on feature selection

and semantic analysis. International Journal of Computer Applications, 146(13)

[12] Maas AL, Daly RE, Pham PT, Huang D, Ng AY, Potts C (2011, June) Learning word vectors for sentiment analysis. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1 (pp. 142-150). Association for Computational Linguistics

[13] MajorLeagueJesus (2018, June 26) Review Counter-Strike: Global Offensive Mautner P, Moucek R (2012) Processing and categorization of Czech written documents using neural networks. Neural Network World 22(1):53–66

[14] Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781

[15] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119)

[16] Pennington J, Socher R, Manning C (2014, October) Glove: global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp.1532-1543)

[17] Plaza L, de Albornoz JC (2012) Sentiment analysis in business intelligence: a survey. In Customer Relationship Management and the Social and Semantic Web: Enabling Cliens Conexus (pp. 231-252).

[18] Pulver A, Lyu S (2017, May) LSTM with working memory. In 2017 International Joint Conference on Neural Networks (IJCNN) (pp. 845-851). IEEE

[19] Saggion H, Funk A (2009) Extracting opinions and facts for business intelligence. RNTI Journal, E (17),119, 146

[20] Seibold M, Jacobs D, Kemper A (2012) Operational business intelligence: processing mixed workloads. It Professional 15(5):16–21

[21] Singh B, Kushwaha N, Vyas OP (2016, November) An interpretation of sentiment analysis for enrichment of business intelligence. In 2016 IEEE Region 10 Conference (TENCON) (pp. 18-23). IEEE

[22] Standard search API – Twitter (2018, June 28) Twitter.

[23] Taboada M, Brooke J, Tofiloski M, Voll K, Stede M (2011) Lexicon-based methods for sentiment analysis.

[24] The Homoeopathic Clinic by Dr. Riyanka Bhardwaj-Homeopathy Doctor in Greater Noida. (2018, August 08)

[25] Fifth International Conference on Advanced Computational Intelligence (ICACI) (pp. 221-223). IEEE