Language Identification for Multilingual Sentiment Examination

Deepali D. Londhe, Aruna Kumari, Emmanuel M.

Abstract: Social media is most popular platform on which users can share their views, reviews and knowledge about various topics, news, products etc. Identifying sentiments or opinions of users is valuable for many e-commerce companies, Hotels, e-learning etc. This opinion analysis is useful for companies to improve their service and products. Due to increase in web users across globe, users happen to post their views freely over the internet. Many different languages are spoken across globe, supporting multilingual nature of social media makes analysis of such text difficult. Sentiment analysis can be conducted using videos, image, text, where text sentiment analysis is most popular form because of freely available contents in the form of blogs, reviews, comments etc. Because of development of social media platform, people can post comment in any language, creates the need for Multilingual sentiment analysis. Sentiment analysis task needs phases such as data collection, pre-processing, sentiment classification and polarity identification. The Multilingual nature needs Script Identification on the input text by labelling the different words used in text along with scripts used to denote them. Various languages used in the text are identified and the Hindi language text written in Romanized script is transliterated to Devanagari script. Text is then completely translated into English language and POS(Parts of Speech) tagging is performed on the obtained text. The aim and purpose of this study is to survey different techniques of multilingual sentiment analysis, and language identification of source text, where n-grams model outperforms all.

Index Terms: Language Processing, Sentiment Analysis, Machine Learning, Lexicon based approach.

I. INTRODUCTION

Substantial spread of technology across the world has increased the number of social media usage among all adults. Due to increase in mobile penetration, users are able to connect to the world of Web easily by using mobile phone applications. People communicate through blogs, forums, wikis, and reviews, facilitating them to share their life events, views, experience, thoughts and knowledge. Such Web-based and mobile based Internet applications allow the creation, access and exchange of user generated contents that is ubiquitously accessible. This produces large volume of subjective information every milliseconds in the form of unstructured text.

 Deepali Londhe, [1]Research Scholar at CSE Department, KL University, Vijaywada, AP, Assistant Professor, IT Department, PICT, Pune, SPPU, Pune, Maharashtra, India.
 Dr. Aruna Kumari, [2]Professor at CSE Department VJIT, Hydrabad, Ph.D.Guide at CSE Department, K.L.University, Vijaywada, A.P.,India.
 Dr.Emmanuel M. Professor at IT Department, PICT, Pune, SPPU, Pune,

Reviews, forums, blogs, micro-blogs, and social network communications help users to coin their opinion, share knowledge and views and influence social, political and economic behaviour worldwide. What the other people think has been an important piece of information for most of us during the decision-making process. We need recommendations before buying a product. Previous to the growth of Internet, we used to rely on advice of friend or relative to recommend product. Due to growth of Web 2.0, user rely upon online advice and recommendation[1]. Opinion of users of product, service etc. value for the companies, as they can analyse what people think about their product, service etc. and target improvements. Many companies have emerged, which help to identify opinions of customers, users, employees of the organization. This can help organizations to grow and strategies in right direction.[3] This natural language form of has to be processed as an NLP problem which is one of the most important components of web mining.

Subjective and objective opinion detection is important challenge in identifying sentiment polarity.[4] People have their opinions and they represent it in many different languages they speak. According to study by KPMG in India and Google April 2017, 'Indian Languages-Defining India's Internet', there will be 75% Indian Language users by 2021 in India's Internet User base. 68% of the users consider local language digital content are more reliable than English. In India Tamil People have highest Internet adoption level, followed by Hindi language users. Limitations of Internet usage in India is limited local language enablement of content and platform[23]. Because of this many users tend to use Romanized English to express their views.

II. LITERATURE SURVEY

Many approaches are identified for Sentiment analysis for English language text. Sentiment analysis of Document or sentence can be done by applying document-level or sentence-level sentiment analysis. For customer/service based applications aspect-level sentiment analysis is also important technique, which identify sentiments based on target aspects[1]. Machine learning approach and lexicon based approach are two techniques majorly used for sentiment analysis on textual data. Machine learning approach uses supervised learning and unsupervised learning methodologies for identifying sentiments. Lexicon based approach uses dictionary based and corpus

based approach for sentiment classification. Many dictionaries are created for identifying sentiment score of given word in Lexicon based approach, such as SentiWordNet., Wordnet etc.[3, 4]. Due to increase in creation of such dictionaries, lexicon based approach is lately very popular, and also gives good accuracy. Machine learning algorithms such as Decision Tree, Support vector machines, Neural Network, Naïve Bayes, Maximum Entropy etc. are used for machine learning based approaches.

Many languages are spoken in the world, but many of sentiment analysis techniques are based on English language[5]. This leads to missing of important information from specific language. To analyze data in many different languages, systems and frameworks are developed. But this lacks the resources available for different languages. Resource rich English language knowledge can be used to improve resources for multilingual sentiment analysis.

Translation:- Due to availability of resources in English, many systems rely on translation of source language to English. Reverse approach i.e. translation of English dictionary/lists into source language can also be deployed. Building resources is brute force and labor-intensive approach for creation of sentiment lexicons for every language.[6]

Sentiment analysis for resource rich English language is not difficult task, as huge labeled data is available for training purpose. Very little literature discusses about multilingual data sentiment classification. Some of the researchers have tried to use different techniques for Sentiment Analysis in languages such as Chinese, Spanish, Hindi, Marathi etc. Mainly two techniques are popular, Machine learning approach and lexicon based approach. Some of them have used Hybrid approach for Sentiment analysis.

Chinese is most widely spoken language in the World. Many researchers have developed Sentiment analysis models for Chinese language text. WU Xing et.al. applied emotion Degree Lexicon by using Cognitive Theories for sentiment analysis[7].They have got average 85% accuracy for Hotel, Book, notebook review dataset. Tao Xu et.al.[8] used S-Hal by identifying semantic orientation of terms for sentiment analysis in Chinese news pages. Wenhao Zhang et.al.[9] have performed Aspect based sentiment analysis for finding product weakness in Chinese reviews, They have got average 82% accuracy for sentence level sentiment analysis. ZHAO Yan-Yan et.al. [10] performed integration of intra and Inter-document evidences for improving sentence level sentiment analysis and got 77.33% and 66.67% accuracy respectively. Yi Hu et.al.[11] performed document level SA on publicly available movie review and a Chinese digital product review corpus, by using two sentence level Topical Term Description Models.

Josef Steinberger et.al.[12]created sentiment dictionaries semi-automatic approach to creating sentiment dictionaries in many languages. Michal Ptaszynski et.al.[13] performed affect analysis in context of characters in narratives such as stories, fairytales, and used Japanese WordNet for affect identification on Japanese online children's books.

Hybrid method for SA to perform automatic classification of French text is incorporated by Sigrid Maurel et al.[14].

Hybrid method is combination of symbolic and statistic, which combined advantages of both methods i.e.robustness of statistical machine learning and manual configuration by symbolic method. Khalid Khalifa et al. [15] used hybrid method combining the results of lexicon-based approach and naïve Bayes classifier for Arabic opinion question answering. Pre-processing of data includes transformation, normalization and tokenization and exploiting auxiliary information (thesaurus). Lexicon based approach used for replacing some words with its synonyms using the domain dictionary. For comparison, three classifiers have been applied which are Naïve Bayes (NB), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). Their experimental results shows that NB outperforms SVM and KNN by achieving 91% accuracy.

Very few efforts are listed for sentiment analysis of more than one language. Nikolaos Malandrakis[16] have created word level affective text analysis model in English and German and also performed the sentence-level news headlines classification using n-gram model by achieving accuracy of 75.9%. Erik Boiy et.al.[17] have used machine learning based model for multilingual web texts sentiment analysis written in English, Dutch and French. They have achieved accuracy results of ca. 83, 70 and 68% for classifying sentiments for English, Dutch and French Web data, respectively, by noting labeling of training samples was tedious task. Ester Boldrini et.al. [18] used EmotiBlog corpus; a collection of blog in 3 languages: Spanish, Englishand Italian for sentiment analysis using different Machine Learning models created by annotation in EmotiBlog and used SVM Classifier for classification.

Hindi is the also one of the most used language on the Internet. 23 official languages are reported in a country like India. Sumitra Pundlik et al.[19]used ontology for classification of Hindi documents into multiple classes. Sentiment analysis was carried out using HindiSentiWordNet (HSWN) to determine polarity of class.For improved accuracy, they have combined HSWN and LM Classifier.

The commonly used Machine Learning classification algorithms for Sentiment Analysis are, Naive Bayes, Max Entropy, Support Vector Machines (SVM).

III. CHALLENGES IN MULTILINGUAL SENTIMENT ANALYSIS

Twitter and Facebook are most popular social networking sites. On these sites people can freely express their opinions in any language of their choice. But there are many challenges to process such data extracted from social media websites. First challenge is extraction of comments. For public contents on websites, crawlers can be designed to retrieve contents/reviews/comments. Facebook and Twitter has provided REST API to extract comments from their sites. Next challenge is preprocessing of these comments. Preprocessing involves many steps depending on the origin of contents. Twitter comments are accompanied with "#" hash tags, mentioning the topic of comment. e.g. #worldcup. This also helps to identify context of the comment. But some hash tags are confusing and may need more

3572

explanation about the topic/context. In sentiment analysis such hash tags can be used to search comments for specific topic. After gathering whole data, the hash tags can be removed from the actual comments.

Data extracted from Face book has many special characters, symbols, emoticons. Many sentiment analysis approaches use these emoticons to contribute in sentiment detection []. Preprocessing of data from Face book also involves filtration of most relevant comments for the specific topic.

One of the most important challenges in social network data is use of slang and abbreviations. Slangs are non-standard vocabulary terms. In order to complete the sentences people tend to use abbreviations of frequently used expressions such as, "OMG" i.e. "Oh my God!" Preprocessing stage need to tackle such data for correctness of comments.

Social networking sites do not imply any grammatical constraints on users to express their views. The comments may not be of correct grammatical form. This can be treated as noise. Syntactical rules can be applied to correct the sentences, phrases and misspelled words.

Multilingual comments are biggest challenge in social networking sites. Comments extracted might contain text from many different languages. A sentiment analysis tool designed specifically for a particular language need to filter out the comments from other languages and keeps the comments from target language. This reduced the size of data, and might hamper accuracy of liner classifier.

IV. LANGUAGE IDENTIFICATION

Multilingual sentiment analysis needs important task of Language Identification. If the source language is identified, the sentiment analysis process can be carried out according to the resources available in the specific language. Good language Identification algorithm decides the success of sentiment analysis task. Natural language processing and machine learning techniques requires data which is annotated with its language. Natural language processing algorithms must be modified according to the grammar the language. Natural language tool kit (NLTK) in Python is most popular natural language processing package for English. Language identification is very useful for various text processing, text mining tasks such as Named entity recognition, Parts of speech tagging, machine translation and multilingual sentiment analysis.

The proposed Language Identification process every token by passing it through all available language models. Language dictionaries are used to classify the tokens for different languages.

A. Lexicon based approach for Language Identification

Dictionary or Vocabulary of languages can be used for identifying language of the text data, this is known as lexicon based approach. It compares tokens from the vocabulary from different languages. If the sentence or document contains many tokens belonging to specific language, then the sentence or document can be classified as particular language sentence or document.[20]

Fig.1 shows the typical flow of the lexicon based language identification process. The language identification process begins by tokenizing the words in input text. The input after pre-processing is tokenized to form language models. The resulting language model contains word lists consists of unique occurrences of words in each language. This generated word list act as spellchecker models which also serve as functional definitions of every language. The testing profile is generated by constructing tokens of the test set into world list. This system then computes a binary matrix of the test profile by searching for each word (of the test profile) in all the training profiles.

The lexicon based language identification model does not tackle miss spelt word. Despite of many words in a particular language, the dictionaries are of limited size. Few words can occur in two languages like Hindi and Marathi; such ambiguity may not be resolved by this method.

B. Phonetic Algorithms for Language Identification

Phonetic algorithms perform fuzzy matching for the phonetic typed data, such as "Swathi" and "Swati". Phonetic methods do matching based on sound produced by the respective word. Most popular algorithms are Soundex algorithm, NYSIIS, Metaphone and double metaphone.[21]

Soundex algorithm

Soundex algorithm assigns same value to similar sounding names, so that they get same value. This is known as soundex encodings. Instead of word search, soundex will search word based on Soundex encoding. Misspelt words can be a problem for database designer. Some words spelling have diversity of ethnic origins, i. e. their pronunciation and spelling are different, and some words can be spelled differently according to different cultures. Soundex helps to resolve such issues. Six phonetic classifications of human speech sounds (bilabial, labiodental, dental, velar, alveolar, and glottal), categorized based on where you put your lips and tongue to make the sounds are considered in Soundex algorithm.

Soundex algorithm reduces the gap between the process of human vocal interaction and fuzzy interaction. It has concise true false processes at the foundations of computer communication. It is an unreliable interface. It is usable in applications which can tolerate High false positives(where words don't match sound of inquiry are returned)and high false negatives(when words that match the sound of the inquiry are NOT returned).

NYSIIS

Developed in 1970 as part of "New York State Identification and Intelligence System(NYSIIS)" algorithm gives better results relatively to the original Soundex using more sophisticated rules for transforming the original word to the result code. This algorithm is designed to work specifically with American names. E.g.WABARLY - Webberley, Wibberley.

Metaphone

Metaphone (1990) algorithm has better efficiency, where for encoding it transforms original word using English pronunciation rules instead of conversation rules. Thus the algorithms loses much less information, since the letters are not divided into groups.

E.g. LWRS -Lowers, Lowerson.

Double Metaphone

Double Metaphone (2000) generate two code values (both up to 4 characters) - one reflects the basic version of word pronunciation, another - an alternative version. It has large number of different rules that account origin of words, focusing on Eastern European, Italian, Chinese and another. E.g. MKTS - Mackiewicz, Mccuthais, Mecozzi.

C. N-grams Approach for Language Identification

N-gram of token can be used for identifying language. N-gram sequence of n items from a given sample of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application. The n-grams typically are collected from a text or speech corpus.[22]

Facebook, Twitter has unwanted advertisements, pictures, hyperlinks, comments referencing using @ etc. This data is removed as it does not help to performs sentiment analysis. After cleaning up above contents the data is then tokenized. Stop words does not contribute to any sentiments and just increase the word count, they are removed in data cleaning phase. Stop words affects the influence of sentiments words thus removal of top words refines the data.

C. Transliteration

The data having Devanagari and Romanized scripts and written in Hindi and English languages are transliterated into Romanized form for simplicity. Transliteration of frequency lexicon of Hindi from Devanagari to Romanized and Devanagari script to Romanized script will be done by this module.
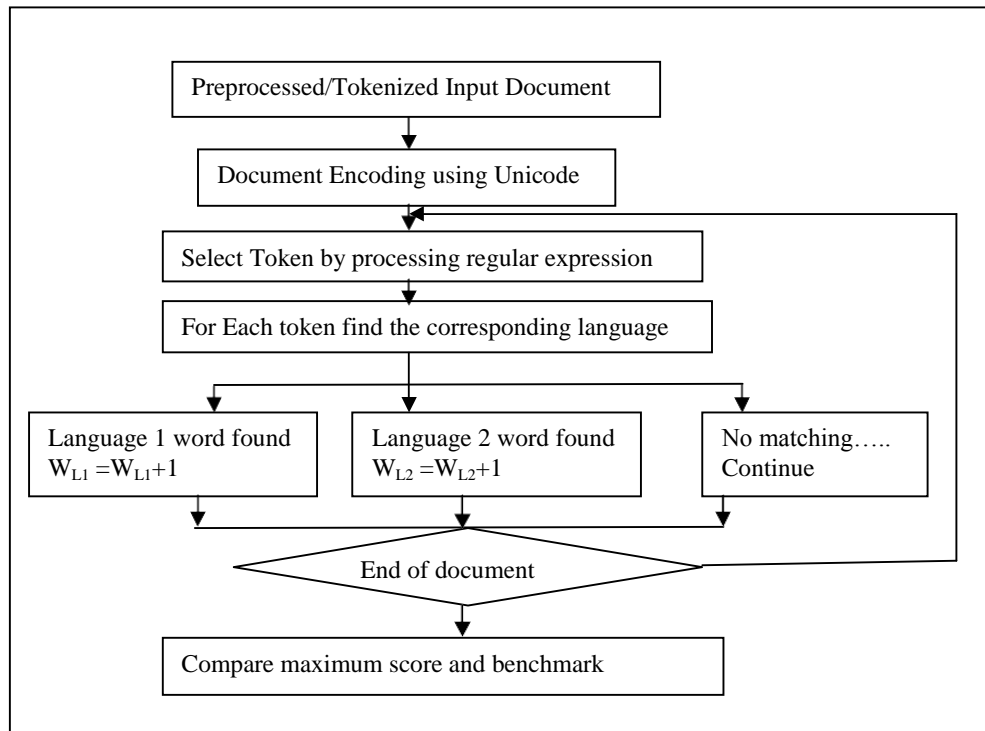

Fig. 1. Lexicon Based Language Identification Process

V. PROPOSED SYSTEM

With reference to Literature survey, three techniques are employed for SA, Machine learning based, Lexicon based, and Hybrid technique. For Multilingual sentiment analysis identification of language is important step. Following are the important steps in proposed architecture.

A. Data collection and pre-processing

A high level model for text analysis includes tasks like text/comments/review extraction by scraping a website, importing it in a specific format, getting it from a database, or accessing it via an API. Text identification process filters interesting text which add noise to the analysis. Such as removal of stop words and other information besides the main content. The data for sentiment analysis is collected by crawling the public pages from Social Networking Sites.

B. Data Cleaning

Data cleaning filters out unwanted text and converts data into structured format. The comments extracted from

D. Language Identification Task

The data considers here is Marathi and hindi text, where there both languages uses Devnagri script. Thus any phonetic algorithm will not perform well for language identification task. Fine grained classification can be done only by comparing n-grams of token for accurate identification of language.

The proposed methodology is based on the assumptions of hidden markov model which in turn uses character n-gram as features. The language model algorithm i.e viterbi algorithm will be used to calculate the most likelihood tags of the words. This algorithm generate character n-gram by applying Viterbi algorithm. Viterbi 6 algorithms helps to calculate various probabilities like start and transition probabilities, which can be further used to tag each word based on its probability in both the languages. Fine grained analysis involves identification of bigrams or trigrams as well.

Language identification module consists of Hindi and English language text in Romanized form. The module will tag each word as Hindi or English. This module will include generation of bigrams from input dataset followed by language identification using transliterated frequency lexicon.

VI. EXPERIMENTS AND RESULTS

A. Lexicon Approach for Language Identification

For this experiment we fetched the english and romanised hindi lexicons which were freely available on the internet, we collected 476,000 hindi words and 150,000 english words. The required time for this approach was much more than other approaches.

But there were problems with this approach. Since one of the problems were that the lexicon approach does not handle variations in the spellings of the words or misspellings. To overcome this problem we used levenshtein algorithm which was again computationally expensive.

B. N-Gram approach for Language Identification

Hidden Markov model is used for identifying language by giving the features like the unigram frequency of the word and the character level n-grams. The viterbi algorithm was used which finds the most likely tags of the words given a sentence based upon the start probabilities and transition probabilities. That means in simple words this approach considers the tag of previous words as an important feature i.e. this algorithm is context based. To train data the unigram words along with their corresponding frequencies calculated on some corpora. The data collected has 50000 unique hindi unigrams and 50000 unique english unigrams with their corresponding frequencies. The character n-grams were generated from these words and the model was trained on those n-grams.

To test the algorithm we collected 2000 mixed code from twitter and manually tagged each word in the sentences and then compared it with the tags tagged by the n-gram algorithm. According to the outcomes shown in above table-I, N-gram algorithm was always more precise than Soundex, Lexicon, and Dmetaphone.

TABLE I
Language Identification Experimental results

Algorithm	Accuracy(%)
Soundex	64.33
Dmetaphone	50.47
Lexicon	74.21
N-gram	90.20

VII. CONCLUSION

The main objective of the proposed work is to identify language for sentiment analysis task. Lexicon based, phoneme based and n-gram based approaches are applied for Language identification task. Lexicon based approach provides 74.21% accuracy, because it doesn't handle variations in the spellings of the words or misspellings. The solution to this problem is to use levenshtein algorithm, which happened to be computationally expensive.

Phonetic algorithms such as Soundex and Dmetaphone provided 64.33% and 50.47% accuracy respectively.

These algorithms have not performed well in this task where English and Marathi text was majorly used. N-gram approach outperformed other algorithms, provided 90.20% accuracy. Hence it is suggested for Multilingual Sentiment analysis, language identification task n-grams can be used for fine grained classification.

REFERENCES

1. Pang, B., Lillian, L., Shivakumar, "V.: Thumbs up? Sentiment classification using machine learning techniques", in: *Proceedings of conference on empirical methods in natural language processing (EMNLP)*, pp. 79–86 (2002)
2. Walaa Medhat , Ahmed Hassan , Hoda Korashy , "Sentiment analysis algorithms and applications: A survey", *in Shams Engineering Journal (2014)* 5, 1093–1113
3. Liu B, "Sentiment analysis and opinion mining", *Claypool Publishers,* May 2012.
4. Kerstin Denecke, "Using SentiWordNet for Multilingual Sentiment Analysis", *ICDE Workshop* 2008.
5. Kia Dashtipour, Soujanya Poria, Amir Hussain, Erik Cambria, Ahmad Y. A. Hawalah, Alexander Gelbukh, Qiang Zhou, "Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques", *Cognitive Computation* (2016) 8:757–771.
6. Stefan Gindl, Arno Scharl, Albert Weichselbraun, "Generic High-Throughput Methods for Multilingual Sentiment Detection", *4th IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST 2010).*
7. WU Xing, LˊU Hai-tao , ZHUO Shao-jian, "Sentiment Analysis for Chinese Text Based on Emotion Degree Lexicon and Cognitive Theories", *Shanghai Jiaotong University and Springer-Verlag Berlin Heidelberg* 2015.
8. Tao Xu, Peng Qinke, Cheng Yinzhao, "Identifying the semantic orientation of terms using S-HAL for sentiment analysis", *Knowledge Based Systems* 2012;35:279–89.
9. Zhang Wenhao, Hua Xu, Wan Wei, "Weakness finder: find product weakness from Chinese reviews by using aspects based sentiment analysis*". Expert System Applications* 2012;39:10283–91.
10. Zhao Yan-Yan, Qin Bing, Liu Ting, "Integrating intra- and interdocument evidences for improving sentence sentiment classification", *Acta Automatica Sinica* 2010;36(October'10).
11. Yi Hu, Li Wenjie, "Document sentiment classification by exploring description model of topical terms", *Computer Speech Lang* 2011;25:386–403.
12. Steinberger Josef, Ebrahim Mohamed, Ehrmann Maud, Hurriyetoglu Ali, Kabadjov Mijail, Lenkova Polina, Steinberger Ralf, Hristo Tanev, Silvia Vázquez , Vanni Zavarella, "Creating sentiment dictionaries via triangulation", *Decision Support Systems* 53 (2012) 689–694.
13. Ptaszynski Michal, Dokoshi Hiroaki, Oyama Satoshi, Rzepka Rafal, Kurihara Masahito, Araki Kenji, Momouchi Yoshio, "Affect analysis in context of characters in narratives", *Expert System Applications* 2013;40:168–76.
14. Sigrid Maurel, Paolo Curtoni and Luca Dini, "A Hybrid Method for Sentiment Analysis", *Statistical Analysis Software (SAS) press*,Grenoble, France, 2008.
15. Khalid Khalifa and Nazlia Omar, Knowledge Technology Group, Center for Artificial Intelligence Technology (CAIT),Faculty of Information Science and Technology, University Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia , "A hybrid method using lexicon-based approach and naive bayes classifier for arabic opinion Question answering", doi:10.3844/jcssp.2014
16. Nikolaos Malandrakis, Student Member, IEEE, Alexandros Potamianos, Senior Member, IEEE, "Distributional Semantic Models for Affective Text Analysis", *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING,* VOL. 21, NO. 11, NOVEMBER 2013.

17. Erik Boiy Æ Marie-Francine Moens, "A machine learning approach to sentiment analysis", *Springer Science+Business Media*, LLC 2008 Inf Retrieval (2009) 12:526–558.

18. Boldrini Ester, Balahur Alexandra, Martı́nez-Barco Patricio, Montoyo Andre´ s, " Using EmotiBlog to annotate and analyse subjectivity in the new textual genres", *Data Mining and Knowledge Discovery* 2012;25:603–34.

19. Prof.Sumitra Pundlik Prachi Kasbekar Gajanan Gaikwad, "Multiclass Classification and Class based Sentiment Analysis for Hindi Language", *Intl. Conference on Advances in Computing, Communications and Informatics* (ICACCI), Sept. 21-24, 2016

20. R. Mahesh K. Sinha, Anil Thakur, "Machine Translation of Bi-lingual Hindi-English (Hinglish) Text" . *In International Conference On IT Convergence and Security,* Jan 2014.

21. Charese Smiley,Sandra Kubler, "Native Language Identification using Phonetic Algorithms" . *In Proceedings of the 12th Workshop on Innovative Use of NLP for Build- ing Educational Applications*, pages 405–412,Copenhagen, Denmark, September 8, 2017.

22. Said Gadri, Abdelouahab Moussaoui, Linda Belabdelouahab-Fernini, "Language Iden- tification: A New Fast Algorithm to Identify the Language of a Text in a Multilingual Corpus" , *In International Conference on Computing and Systems (ICMCS)*, Marrakech, Morocc, 14-16 April 2014.

23. 'Indian Languages-Defining India's Internet', a study by KPMG in India and Google April 2017. Available at https://assets.kpmg/content/dam/kpmg/in/pdf/2017/04/Indian-languages- Defining-Indias-Internet.pdf

AUTHORS PROFILE

Deepali Dashrath Londhe is Masters in Computer Science and Engineering. Currently perusing Ph.D.from K.L.University, Vijaywada, A.P., India. Her areas of interest are Information Retrieval, Data mining, Human Computer Interaction, Software Modeling and Design, Data Science and Big Data analytics, and Internet of Things etc. She has published seven papers in International Journals and 3 papers in International conferences. She has contributed by writing chapters in Book on Data Science and Big Data analytics, Software Modeling and Design. She is currently working as Assistant Professor in Information Technology department, at Pune Institute of Computer Technology, Pune, Maharashtra.

Dr.Aruna Kumari, Professor, at CSE Department VJIT, Hyderabad. She has more than 70 research articles in International Journals and Conferences, She is Young Scientist awardee and currently working on Research Project, "Design and Development of Effective privacy preserving data mining for cardiac, cancer and diabetic healthcare", funded by DST-SERB. She has organized several national Conferences/Workshops. She is Ph.D. Supervisor at CSE Department, K.L.University, Vijaywada, A.P.,India

Dr.Emmanuel M is a Professor in Department of Information Technology at Pune Institute of Computer Technology, Pune, Maharastra, India. From the last 17 years he has been involved in teaching the under graduate and post graduate students. He held the administrative position as Head of the Department of Information Technology Department, PICT Engineering college for 5.6 years. He published 55 papers in National/International Journals & Conferences. He is a member of several Editorial & Review Boards of International Journals and also member of several professional bodies. He is Member of Board of Studies of Information Technology branch, SPPU, Pune. His research interests include Data Mining,, Information Retrieval, Big data, Cloud Computing and Medical Image Processing. He can be reached at emman2001@gmail.com.