Sentiment Analysis of Transliterated Hindi and Marathi Script

Mr.Mohammed Arshad Ansari* and Prof. Sharvari Govilkar**
*Department of Information Technology, PIIT
mansari@student.mes.ac.in
** Department of Information Technology, PIIT
sgovilkar@mes.ac.in

Abstract: There is a growing research on sentiment analysis of various languages, which is being supplanted heavily by those same techniques and methods being applied on the mix code or transliterated text for the same purpose. This growing research is a result of necessity created through the advent of social media as well as textual analysis of the data being collected online. This paper, rather than being a pioneer, is about extending that research for further improvement. Herein, we assess the existing status, standards and achievements of the researchers in the given field and supplant it without proposed methodology to increase precision. Although, the current work is a proposal with improvements over established techniques, it is also however going to be quite comparative when it comes to the existing findings. The idea is to not just improve what has already been built or shown to be true, but also check if the simplest approach is still the best way to proceed or not. By this we mean the existing direct supervised learning for sentiment analysis, without much NLP or language specific work. Since we shall be testing our approach against the existing state of the art as well as entering the area previously not under coverage (Marathi transliterated text), this work is bound to make great strides in the field of sentiment analysis.

Keywords: Mixed code script, English, Hindi, Marathi, Transliteration, Sentiment Analysis.

Introduction

Sentiment analysis is a process of analyzing natural language and figuring out the sentiments involved or expressed through the source material, with respect to the topic. The basic idea behind sentiment analysis is that each textual sentence may or may not contain some kind of polarity, expressing a degree of emotions along with the information. It is much easier to read in to those polarities when the text is spoken and not written due to the tone of the speaker; whereas, in case of written text, it is the context that is useful while determining the polarities in the statements. Sentiment analysis has grown to be one of the most important research areas when it comes to textual analysis on the web. Reason being, obviously, is to be able to make sense of the data as well as to understand the tone of information being provided. There are numerous applications, ranging from product/customer support review to improve quality of service (QOS) by corporations to understanding geo-political motivations when certain news breaks. People react on social media, especially when they are charged emotionally and when emotions take the form of textual content to vent, it has been observed that it does in a manner which is more close to a person's mother tongue.

Hindi is spoken by more than 500 million people around the world, making it one of the most spoken language in the world. Besides, English has turned out to be an international language, a lot of people speak English on the internet, however; as described above, there are instances when people use English language to phonetize and express in a foreign language. This is seen far more in India subcontinent, where people prefer to write using English alphabets, but most often, use the words from the mother tongue. If we only look at all the YouTube comments (especially if they are about some controversial issues), we would see a lot of usage of such transliterated messages or mix-script writing. Another behavior worth noting is related to vocabulary. People from subcontinent use words such as 'Bye', 'Thank you', 'Good night', 'Please', 'Sorry' and intermix them with their native tongues. This mixture of language has been observed profoundly at varying levels of society. Therefore, it would not be very far-fetched to say that the languages are evolving by mixing language themselves. This forms the necessary reason for why there needs to be analysis of mixed-languages and it starts with analyzing that which is mostly available, the mixed-script. Here, we are not going to invent something new, nor are we going to do something entirely differently. However, the purpose behind this work is to stand on the shoulders of giants and take the research of what has already been done to what it can be. This we strive to do, by improving the performances by innovatively applying techniques which have worked better in other cases. Therefore, as it will be seen, our proposed approach as well is a mixture of disparate attempts in varying domains (even slightly) to come together for better whole.

Sentiment analysis is a lot tougher for languages that are outside Eurozone, due to their lexical syntax being very different from European languages as well as due to majorly, less amount of work being done on it. Semantic analysis requires

annotated text corpus to train classifiers, which is most of the time a very huge manual task. It has been undertaken for English and for many other European languages, while at the same time, work from one supplementing work for another language, due to the similarities existent in those languages. When it comes to languages such as Hindi and Marathi, such resources are very less compared to the above mentioned languages. More so for Marathi, since a lot of work has been done and progress made in case of Hindi. Most of the reason for the under development of the research for these languages are (1) Not much annotated textual corpus needed for training, (2) Lack of basic language tools like taggers and parsers. These problems will be solved in time and this work is a part of all the woks which will finally solve this problem. Having expressed the problem, in this work, we also laude the work that has already gone in to this respective field, without which this would not have been possible. It is really interesting to note, that a great amount of effort has just started pouring in for this particular part of sentiment analysis. It is naught with great anticipation that this work is being progressed. Besides, as the sentiment analysis of the textual data being to shape more and more, the greater the benefit will be to the field of general AI. When it comes to human capacity, not representing emotions would be the biggest gap in the domain, which exactly is sentiment analysis has started to fill.

Motivations

We can discuss two kinds of motivation herein, one being in general about sentiment analysis and the other a bit specific about this very work itself. Sentiment analysis has tons of applications, especially in the current era of social media. The entire planet's population is connecting with one another and learning about each other's cultures and assimilating ideas from one another and then followed by sharing ideas, concerns and assaults on the social media. Whatever may be the case, all the information being transferred is textual in nature. It becomes of vital import that the sentimental exchanges happening on such channel is being monitored. For example, twitter and Facebook led to the entire Arab world to be engulfed in flames. The previous example was just to elaborate how the social media and in extension written media, with the emotional content, can literally change the world. Having understood the important, let's look at many possible applications for sentiment analysis.

- Product / Service Review; Product here can extend from movies, daily usage products to books, etc. They are usually reviewed either on social media or sites dedicated to such reviews. Examples being Amazon, Google/IPhone App store, Good Reads, etc. These reviews allow other users to decide whether they want to buy reviewed product or not. Similarly, service providers like ISP, Telecom providers, etc. are interested in review of the service they are providing, including customer service reviews. These reviews help the providers to improve the service by addressing the negative aspects and focusing more on positive ones.
- Discourse Analysis on topics that range from philosophical to wars between countries are also candidates for such analysis. It becomes really important when taking decisions whether the debate pertaining to such decisions are emanating from emotions or logically grounded arguments.
- Feedback Analysis for teachers from students or about government from population. They all have one thing in common, that is, they are all textual and by extension is candidate for sentiment analysis.
- Other areas such as emails analysis, twitter/Facebook feed analysis or blog analysis, help in understanding the emotions of authors regarding the topics, which help focus on problems, which can be the cause of negative emotions.

Above being the motivation for sentiment analysis in general, let's consider the motivation for this specific work. As explained in introduction as well as in the above listed areas of interest; where the source is always textual data, this data usually consists of mixed script in terms of language. That being the what, let's look at why. And more specifically why Hindi and Marathi. Marathi by itself is playing catch with Hindi, where Hindi language is making strides in this research. Although there aren't many Marathi speakers in comparison to Hindi itself, but it is still spoken by millions of people in the region of Maharashtra, which by the way has a lot of literature that has yet to be digitized and benefit the world with it. The real reason is the aspect of completion. Many terms like 'Layi Bhari' and many other slangs have crept from Marathi to Hindi and then taken to the country as whole through Bollywood movies. Many inside jokes in many Bollywood movies have their roots in Marathi language and regional aspects. These will never be easily covered if Marathi itself doesn't become partial focus of the research itself. Although, there are very few input sources to consider for Marathi, there still exists some, in form of YouTube comments, etc., which can be part of this research. Going back to Hindi itself, a lot of textual resource considers mixed script statements as noise, which definitely contains gold from sentiment analysis perspective. And, therefore, we have decided to augment the existing research by improving the precision where research is being performed and pave the way where the research is still lagging behind.

Related Work

Code Mixing, Language Identification and Transliteration

Code mixing has been done for more than a couple decades and was investigated during initial period by Gold [1] for the purpose of language identification. The same phenomenon for Indian languages was worked upon by Annamalai [2], pioneering the research field for the subcontinent languages. Recently, it was investigated by Elfarti [3] and was termed as linguistic code switching by the research group. Karimi [4] made the case for machine translation for the purpose of transliteration in the survey and suggested transliteration based on phoneme based approach and transliteration generation using bilingual corpus, while presenting the key issues that arise during the transliteration process. Dewaele [5] pointed out the strong emotional presence as being the main marker for the existence of code switch that happens in textual corpus. Gupta et. al. [6] mined the transliteration pairs between Hindi and English from the music lyrics of Bollywood songs for Fire'14 shared task, which is quite handy for training in language sentiments. Deepti Bhalla et.al. [7] have given a rule based approach to perform transliteration from English to Punjabi language through machine transliteration. Although, the focus is primarily on named entity.

The issue of identification of language of the code - mix script is another challenge that has been answered by the research community. A statistical approach was proposed by Kundu and Chandra et. al. [8] for the automatic detection of English words in Bengali + English (Benglish) text. A conditional random field model for weakly supervised learning model was used for word/token labelling by King and Abney [9] with a good result of > 90\%. Barman [10] used Facebook user data for identifying the language in mixed script and concluded that the supervised learning outperforms the dictionary based approaches. POS Tagging and transliteration efforts for Hindi + English data on social media was experimented upon by Vyas et. al. [11] and came to the conclusion that any operation on transliteration text will largely benefit from POS tagging. Another approach was given by Kulkarni et al. [12] to use genetic algorithm for identification of Marathi and Sanskrit words. In another work by Jhamtami [13], it was shown that the technique which use POS tags of adjoining words along with char sequence to achieve F1 score of 98%.

Sentiment Analysis

Although sentiment analysis is being worked upon for quite some time now and it has already entered the mainstream application. There are works being done for the transliteration of Indian languages, out of which some have been covered in this section. A survey of sentiment classification was performed by Pandey [14] covering the techniques being used for Indian languages. Joshi et.al. [15] performed experiment to compare three approaches for the sentiment analysis of Hindi text and found that HSWN performs better than Machine Translation approach but under performs in language training of sentiment corpus in Hindi. This was, however; performed in 2010 and the HSWN has been continually improving past these experiments. The same result was reiterated with by Balamurali et. al. [16]. Kashyap [17] found a way to perform Hindi Word Sense Disambiguation using WordNet with encouraging results for nouns. Subjective lexical resource was developed by Bakliwal et. al. [18] by using only WordNet and graph traversal algorithm for adverbs and adjectives.

Balamurali A R et. al. [19] performed experiment to figure out in language supervised training of sentiments against the machine translated source for sentiment analysis. They found that the MT based approach under performs much worse compared to in language training of sentiment. Fuzzy Logic membership function was used to determine the degree of polarity of the sentiment for a given POS tagged preposition by Rana [20]. Hindi SentiWordNet was developed by Balamurali A. R. et. al. [19] using the SentiWordNet by using linked WordNet. HSWN along with negation discourse was applied by Pandey [21] and Mittal et. al. [22] or sentiment analysis of Hindi language text corpora, with the accuracy of 80.21 achieved. Work on multi language sentiment analysis on twitter feeds of Sweden politicians was undertaken by Lucas Bronnimann [23], which heavily depended on the use of emoticons for emotional polarity detection. S Tembhurnikar [24] used LDA for sentiment classification on mixed tweets, however; filtered out the non-English words from the tweets before the analysis. Jagmeet Singh [25] used twitter feeds API for performing sentiment analysis in mixed languages with techniques such as dictionaries look up, taxonomy and ontology based analysis. There is only one work done on the sentiment analysis of Hindi transliteration by Srinivas [26], [27] and the approach taken was to tag words with identified language and then run against respective POS tagger for languages and sentiment analysis done on the output. The approach yielded 85\% precision.

Proposed Approach

The current word considers Hindi/Marathi text in Romanized script as input which may contain phonetic words, sounds; however, it is not considering social language like gr8, rt, f9, etc. as input source. For now, it is considered as noise for the result of this work. Although, we are not performing sentiment analysis on English text as part of this text, which is simply because it has been under taken in many works preceding this one. Therefore, concentration will solely be on the text which is transliterated Hindi or Marathi. Also, the architecture of proposed approach has integration in mind and hence, it will be

able to plug social sentiment analysis or twitter sentiment analysis or plain English analysis and will work only for the transliterated text, while taking inputs from the mentioned analyzers for their established polarity. The results can be merged and shown to have improved the overall accuracy.

System Architecture

There are going to me multiple approach for testing to be implemented as part of this work. The purpose of those work will be to ensure that the proposed system performs better than what has been accomplished by other researchers. Although, here we will only go into the actual proposed system to understand its working and predict the possible improvements.

The proposed approach we are to take in this work comprises of extending the work of Srinivas [26] with multiple improvement points at multiple level of the process. Each step is listed below with the improvement suggested from this work. All the work done on that paper has been uploaded to the website [28], which we shall be using in this paper extensively and building on top of it.

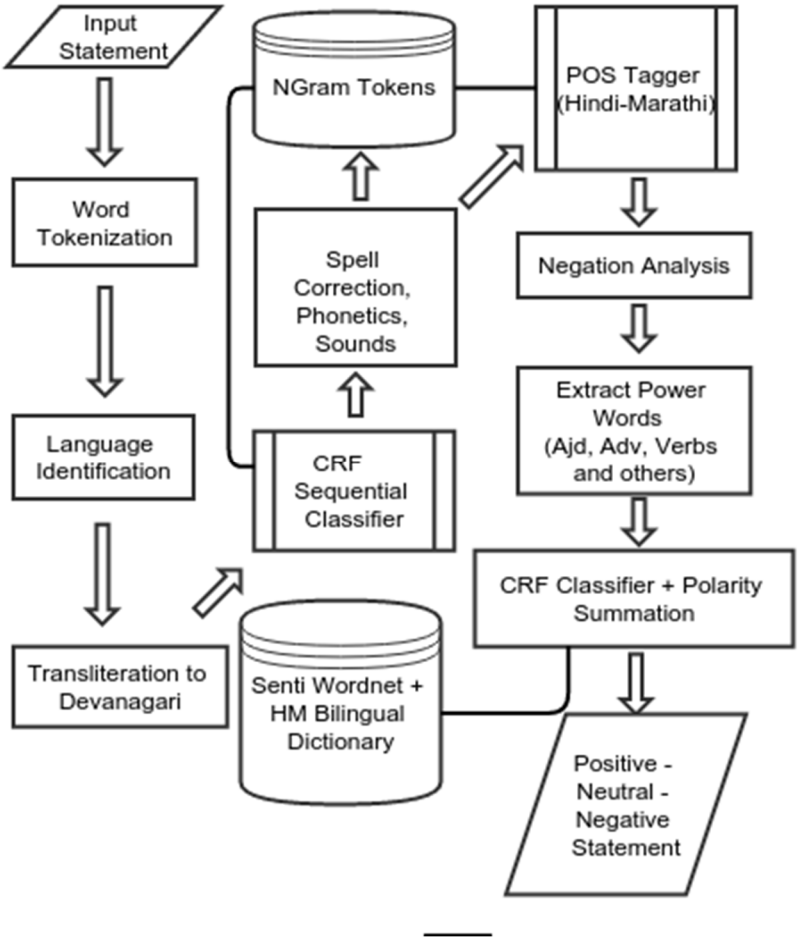

Figure 1. Process flow diagram

Text Normalization

Text normalization step has been covered heavily in work of Srinivas [27], which has following steps:

- Language Identification; Tagging of the words as <word>|<tag>, where <tag> can be E or English, H for Hindi and M or Marathi,
- Spelling Corrections; There are multiple ways to write Mujhe in Hindi such as muze, muje, etc. To come to common and widely used spelling becomes very important. [29]
- Ambiguous words; Words such as 'me' means same thing in English and Marathi, where as in Hindi it is sometimes used to say inside with another spelling being 'mein'. [17] [13]
- Sounds; Words such as aww, oohh, ouch, ewww, etc. They do contain rich information when it comes to sentiments.

- Phonetic words; Words such as plees usually is misspelling of word please, spoken in some areas of the subcontinent. It gets written too in the similar manner.[30]
- Transliteration; Conversion of Hindi/Marathi words written in English to appropriate Devanagari script.

All then above enumerated steps have been covered by Srinivas [27] and doesn't require us to go in the details of those, however; we will look at some of those steps to get a proper grip on the subject. At this point, this work will simply reuse those steps for Hindi and try to closely perform them for Marathi as well. [31]

Work-Token Normalization

The process here is really simple to explain, but quite interesting to develop. This is a required step as a sort of preprocessor, to enable the words to be converted appropriately to their respective languages. Words like "aww", "reaaaly", etc. needs to be normalized and the techniques to do are covered by Srinivas [26], [27] both. These methods are already tested with some accuracy in the above mentioned papers themselves.

Language Identification Tagging

The first step is to tag the language identifier for every word-token, using the techniques described in Kundu and Chandra et. al. [8] and King and Abney [9]. The output of this step will be more or less like in example given below:

"Yeh acchha din hai. Let's go now"
"Yeh|H acchha|H din|H hai|H. Let's|E go|E now|E".

Table 1. Language Identification Algorithm

Algorithm: Language Identification Description: Works only with bilingual settings. Here other language is could either be Hindi or Marathi
Input: Sentence Output: Tagged Sentence Given: English Dictionary Slang Dictionary [Besides slang words, it also contains emotional expressions] Other Dictionary [Hindi / Marathi]
Steps: For every word [u] in Sentence: 1. If word in slang: a. Y = give S[u] probability of 0.7 b. N = give S[u] probability of 0 2. If word [u] passes the phonetic emotions expression test, then convert it to normal spelling and give S[u] probability of 0.9 3. if word exists in Transliterated Other Language Dictionary a. Y = give word O[u] probability of 0.7, E[u] probability of 0.3, S[i] probability of S[u] * 0.3 b. N = give word O[u] probability of 0.5, E[u] probability of 0.5 4. Transliterated the word and check if word exists in Other Language Dictionary a. Yes = give word O[u] probability of 0.7 * O[u] and E[u] probability of 0.3 * E[u], S[u] probability of S[u] * 0.3 b. No = give word O[u] probability of 0.3 * O[u] and E[u] probability of 0.7 * E[u] 5. If word exists in English Dictionary: a. Yes = give word O[u] probability of 0.7 * O[u] and E[u] probability of 0.3 * E[u], S[u] probability of S[u] * 0.3 b. No = give word O[u] probability of 0.3 * O[u] and E[u] probability of 0.7 * E[u] 6. If S[u] > O[u] and S[u] > E[u]: a. Identify Word[u] as Slang 7. Else If O[u] < E[u]: a. Identify Word [u] as English 8. Else: a. Identify Word [u] as Other [English / Marathi] Return Sentence updated with identified language

To this effect, same approach can be accomplished for Marathi text. Once we have tagged all the word-tokens with their corresponding language identifier, we can move to next step. However, there are going to be ambiguous words "ho", "me",

etc., which shall use semi supervised learning for handing based on context. The plan is to also test again HMM models to discover the accuracy difference.

Language identification will be based on the approach of using hidden markov models trained on the n-gram generated from corpus to be able to produce probabilities for each work-token when considered with its neighboring words to identify the language to which the token belongs.

In all the related work, there was a need for transliteration mechanism in play on the fly. The reason for it being the de facto method of choice is because it allows the usage of POS tagging to work with the text, which would only work on respective script for a given language. This database will be trained using Hindi - English transliteration pairs collected from Fire 2013 found at [28] as well as result of another previous work by Gupta et. al.[6]. This trained model will then be used to convert all the words in Hindi WordNet to ensure greater coverage of incoming input word tokens. In case of Marathi, a similar thing will be done and associated with Hindi WordNet through Hindi -Marathi bilingual dictionary.

POS Tagging, Discourse analysis, SentiWordNet

Before sentiment analysis can be performed, it is necessary to deal with few important things. We are striving to extend and improve upon earlier work such as Srinivas [27] and therefore following much in the same footsteps. Both the steps are explained further below. Once we have the document in English or Hindi, the next step is to run it through POS tagger based on respective language. The approach will be straight forward as detailed here [11]. The POS Tagged prepositions then shall be run through the negation discourse analysis to invert the POS tagged adjectives and adverbs in case of negative discourse as explained by Pandey [21] and Mittal et. al. [22].

Table 2. Polarity Identification Algorithm

Algorithm: Polarity Identification
Description: Works only with bilingual settings. Here other language is either be Hindi or Marathi

Input: Sentence [Language Tagged]
Output: Sentence Polarity P = {} for each index j to hold the polarity of each word
Given:
POS Tagger

Steps:

1. Convert Sentence into language based phrases for each word collection belonging to same language as a phrase
2. For every phrase [u] in Sentence:
    a. If the root in phrase is negation then mark Negation = True, otherwise False
    b. POS Tag the phrase [u] using the corresponding language POS Tagger
    c. Identify Named Entity in the phrase [u]
    d. For every word [j] in Phrase [u]:
        1. if word [j] is named entity:
           continue
        2. if word [j] is tagged as slang or English:
           P [j] is polarity of slang in English SentiWordNet
        3. If word is tagged Hindi:
           P [j] is polarity of the word [j] in HSWN
        4. If word is tagged Marathi
            a. Get approximate meaning* of word from Hindi – Marathi bilingual dictionary as jnew
            b. P [j] is polarity of word [jnew] in HSWN
        5. If Negation is true, then inverse the polarity of all words in the phrase by multiplying it with -1
3. Return P containing each word with its corresponding polarity

Return Sentence updated with identified language

The output of POS Tagger shall be used to look up senti-word identifier for the word groups using SentiWordNet or HSWN, for English and Hindi, respectively. HSWN has been improved by Pandey [21] by making additions to it and that will be used in this work. Here, there are three major improvements we are considering. Since, it was established [26] that the basis for sentiment analysis being POS tagged adjectives and adverbs gives much better result that depending directly on lexicon or WordNet look up for each work, we would be going that route. Secondly, addition of discourse analysis would further enhance on the existing work [27].

Sentiment classification using classifier

Once we get sentiword identifier for each token - word, next step is to put it through the classifier which will give the polarity of the statement provided. This polarity checking decision can be as simple as simple summation of all word-token sentiment polarities or further analysis can be performed to figure out what really is the polarity of word-token and its membership with negative, positive or neutral. This step being the vital one can be accomplished using most trust classifier like SVM, Random Forests, however; impetus shall be given on naive Bayes classifier for brevity's sake.

Example Flow

'Kitni der se ticket cancel nahi ho rahi hai' becomes kitni|H= कितनी der|H= देर se= से|H ticket|E cancel|E nahi|H= नहीं ho|H= हो rahi|H= रही hai|H= है

Table 3. Word POS Tagged

कितनी	Adjective	QF
देर	Adverb	NN
से	Verb	PSP
Ticket	Unk	JJ
Cancel	Unk	NN
नहीं	Adverb	NEG
हो	Verb	VM
रही	Verb	VAUX
है	Verb	VAUX

Negation discourse analysis and polarity extraction example

नहीं and हो are closely associated with one another. Negation discourse analysis [21] works on the subtree level, which in this case is post the word हो. So all the words following 'nahi' will be part of its subtree. Hence, all the polarity from that point onwards will be reverse.

polarity = कितनी|adj=INC + देर|adv=NEG + से|v=NEU + ticket|NN=Neu + cancel|=NEG + नहीं|adv=NEG + हो|v=NEU रही|v=NEU + है|v=NEU

polarity = (INC * NEG) + NEU + NEG + NEG REVERSE (NEU + NEU + NEU) = (2 * -1) + 0 + -1 + (-1 * (0 + 0 + 0) = -3

Quite clearly, we have input as POS tagged statements with greater emphasis on adjectives and adverbs that are inverted in case of negation present in the preposition. Once we have this tagged information, we would like to test on both the process of simple polarity count summation of the given input and training the classifier, in order to come up with the best possible result in terms of accuracy.

Conclusion

The most important aspect of this work i.e. the results are what is coming next. We will show that the approach proposed in this work performs better than all the work presented here in literature, when considered independently. It is the synergy, which the approach presented there, promises. The implementation will happen for all ways that differ from the approach too, so that comparisons can be made and conclusions drawn without the strawman arguments.

There is a lot of work to be performed before any concrete conclusion can be expressed, however; There is a great possibility that the approach suggested in the given work will result in improvement in the field of sentiment analysis, that can again be extended for greater language coverage in as well as out of Indian languages. These strides towards such improvements will result in machine's being able to understand human sentiments better, which is one of the greatest challenge being faced by the research in general AI. Ours is but a small step towards that goal. It will not be too farfetched to believe that the improvements will range from 5 to 10 percent improvement where we will see the accuracy reach 95 percent.

The authors wish to thank all the peers who helped enable this work to come to fruition.

References

[1] E. M. Gold and T. R. Corporation, “Language identification in the limit,” *Inf. Control*, vol. 10, no. 5, pp. 447–474, May 1967.

[2] E. Annamalai, “The anglicized Indian languages: A case of code mixing,” *Int. J. Dravidian Linguist.*, vol. 7, no. 2, pp. 239–247, 1978.

[3] H. Elfardy and M. T. Diab, "Token Level Identification of Linguistic Code Switching.," in *COLING (Posters)*, 2012, pp. 287–296.
[4] S. Karimi, F. Scholer, and A. Turpin, "Machine transliteration survey," *ACM Comput. Surv.*, vol. 43, no. 3, pp. 1–46, Apr. 2011.
[5] J.-M. Dewaele, "Emotions in Multiple Languages." 2010.
[6] K. Gupta, M. Choudhury, and K. Bali, "Mining Hindi-English Transliteration Pairs from Online Hindi Lyrics," *Proc. Eighth Int. Conf. Lang. Resour. Eval.*, pp. 2459–2465, 2012.
[7] D. Bhalla, N. Joshi, and I. Mathur, "Rule Based Transliteration Scheme For English To Punjabi," *Int. J. Nat. Lang. Comput.*, vol. 2, no. 2, pp. 67–73, 2013.
[8] B. Kundu and S. Chandra, "Automatic detection of English words in Benglish text: A statistical approach," in *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, 2012, pp. 1–4.
[9] B. King and S. Abney, "Labeling the Languages of Words in Mixed-Language Documents using Weakly Supervised Methods," *Proc. NAACL-HLT*, no. June, pp. 1110–1119, 2013.
[10] U. Barman, A. Das, J. Wagner, and J. Foster, "Code Mixing: A Challenge for Language Identification in the Language of Social Media," in *First Workshop on Computational Approaches to Code Switching*, 2014, pp. 21–31.
[11] Y. Vyas, S. Gella, J. Sharma, K. Bali, and M. Choudhury, "POS Tagging of English-Hindi Code-Mixed Social Media Content," pp. 974–979.
[12] P. P. Kulkarni, S. Patil, and G. Dhanokar, "Marathi And Sanskrit Word Identification By Using Genetic Algorithm," vol. 2, no. 12, pp. 4588–4598, 2015.
[13] H. Jhamtani, "Word-level Language Identification in Bi-lingual Code-switched Texts," pp. 348–357, 2014.
[14] P. Pandey and S. Govilkar, "A Survey of Sentiment Classification Techniques," *Ijser*, vol. 3, no. 3, pp. 1–6, 2015.
[15] A. Joshi, B. A. R, and P. Bhattacharyya, "A Fall-back Strategy for Sentiment Analysis in Hindi: a Case Study," no. October 2015, 2010.
[16] a R. Balamurali, A. Joshi, and P. Bhattacharyya, "Harnessing WordNet Senses for Supervised Sentiment Classification," *Proc. Conf. Empir. Methods Nat. Lang. Process.*, no. 2002, pp. 1081– 1091, 2011.
[17] M. Sinha, M. K. R. R, P. Bhattacharyya, P. Pandey, and L. Kashyap, "Hindi Word Sense Disambiguation."
[18] P. A. V. V. Akshat Bakliwal, "Hindi subjective lexicon: A lexical resource for hindi adjective polarity classification," *In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), Chair.* .
[19] A. Balamurali, "Cross-lingual sentiment analysis for Indian languages using linked wordnets," *Proc. COLING 2012*, vol. 1, no. December 2012, pp. 73–82, 2012.
[20] S. Rana, "Sentiment Analysis for Hindi Text using Fuzzy Logic," *Indian J. Appl. Res.*, vol. 4, no. 8, p. 16, 2014.
[21] P. Pandey and S. Govilkar, "A Framework for Sentiment Analysis in Hindi using HSWN," vol. 119, no. 19, pp. 23–26, 2015.
[22] N. Mittal and B. Agarwal, "Sentiment Analysis of Hindi Review based on Negation and Discourse Relation," in *Sixth International Joint Conference on Natural Language Processing*, 2013, pp. 57–62.
[23] L. Brönnimann, "Multilanguage sentiment - analysis of Twitter data on the example of Swiss politicians," 2013.
[24] S. D. Tembhurnikar, "Sentiment Analysis using LDA on Product Reviews : A Survey," no. Ncacc, pp. 22–24, 2015.
[25] J. Singh and K. Mahajan, "SENTIMENTAL ANALYSIS IN MASHUP LANGUAGES," vol. 8, no. 4, pp. 663–667, 2015.
[26] S. Sharma, P. Srinivas, and R. C. Balabantaray, "Sentiment analysis of code - mix script," in *2015 International Conference on Computing and Network Communications (CoCoNet)*, 2015, pp. 530–534.
[27] S. Sharma, P. Srinivas, and R. C. Balabantaray, "Text normalization of code mix and sentiment analysis," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2015, pp. 1468–1473.
[28] S. Sharma and P. Y. K. L. Srinivas, "Linguistic Resource Website," 2015. [Online]. Available: http://linguisticresources.weebly.com/downloads.html.
[29] E. Clark and K. Araki, "Text normalization in social media: Progress, problems and applications for a pre-processing system of casual English," *Procedia - Soc. Behav. Sci.*, vol. 27, no. Pacling, pp. 2–11, 2011.
[30] D. S. Rawat, "Survey on Machine Translation Approaches used in India," no. 6, pp. 36–39, 2015.
[31] D. Pisharoty, P. Sidhaye, H. Utpat, S. Wandkar, and R. Sugandhi, "Extending Capabilities of English to Marathi Machine Translator."