Early Detection of Depression in Multi-lingual Code-Switching

Brij Rokad , **Dhaval Bhanderi** , **Rutu Gandhi** , **Shrinidhi Adke**
University of Georgia
Athens, GA, United States
{brijrokad, dhaval.bhanderi, rutu.gandhi,shrinidhi.adke}@uga.edu

Abstract

Detection of depression - one of the most occurring mental illnesses - is a crucial task, especially in early stages. Previous works had demonstrated the use of Natural Language Processing(NLP) to help detect depression from monolingual texts. In the case of multilingual texts, code-switching is a challenge for NLP tasks. While some work has been done on code-mixed text concerning sentiment analysis and emotion prediction, our work aims to apply NLP techniques to identify depression in code-switched social-media texts. Here, we have analyzed this problem and presented a Hindi-English code-switched corpus extracted from twitter by manually labeling with the associated causal language of depression. The corpus was analyzed to study the psycho-linguistic aspect of expressing depression. Further, we propose various classification models that use supervised machine learning techniques for detecting whether the person writing the text is expressing depression or not. We make use of multiple features extracted from the available text and also studied the effect of transfer learning for classification. We observed that using a pre-trained model on different code-switched data helped us learning the features and improved the classification accuracy to 94%.

[1] [2] [3]

[1] CSCI 6900: Natural Language Processing(2019)
[2] LING 6570: Natural Language Processing(2019)
[3] The code of this project is available at https://github.com/dvlbhanderi/Depression-Detection-In-Multilingual-Code-Switching

1 Introduction

Researchers have shown that applications of Natural Language Processing can be highly useful in the detection of health illnesses, especially in mental illnesses and early detection of depression. According to the Centers for Disease Control and Prevention (CDC), depression is one of the most common and widely spread mental illness. Depression is also classified as a mood disorder or emotional disorder. Emotions are one of the major factors in the early detection of depression. The challenging part is that people express different emotions in different ways, but that availability is not an issue here, due to the popularity of opinion-rich resources such as online review sites, forums, and micro-blog websites emotion analysis in the text has become highly significant in obtaining useful information for studies on social media. So far most of the research has focused solely on analyzing depression in monolingual text.

However, code-switching conversations or texts are commonly found in the multilingual community and social media. Emotions expressed in these texts are presented in both monolingual and bilingual ways. [S1-S4] are four examples of code-switched tweets and reviews that contain both Hindi and English words. While [S1] expresses the neutral emotion in English, but if we look for Hindi context around the English word and translate it in Hindi, then it is quite negative, whereas [S2] expresses the negative emotion in both Hindi and English. But the example three [S3] is a positive review, where it shows a good characteristic of Karan Johar towards his subjects. Example four [S4] can be classified as slightly negative, it seems that people were disappointed after the interval. These types of emotions such as disappointment, sadness or anxiousness can also lead to depression. From [S1-S4], we can see that it is much more difficult to detect depression in the code-switching text than in the monolingual ones since depression in code-switching text can be expressed in either one, two or many languages. Consider [S2] where the English Text “waste” conveys the negative emotion, which is indeed triggered by the surrounding text in Hindi. Moreover, some emotions are expressed in two languages simultaneously, statistics show (Wang et al., 2017) that 14.2 percentage of text contains more than one emotion.

S1: माहि का तो band बज जाएगा अभी
T1: Mahi will get f***ed up now.

S2:: साला यह जनम तो Waste ही हो गया मेरा
T2: Shit, my birth has gone to waste.

S3:: करण जौहर अपने subject के प्रति honest हैं
T3: Karan Johar is honest towards his subjects.

S4: Interval के बाद film उलझ जाती है
T4: After interval the movie is too confusing.

In code-switching text, depression could be expressed in more than one language. Different emotions within a single text can be closely related to each other. Therefore, the major challenges of early detection of depression in code-switching texts are to explore bilingual information in each text and correlate different emotions within a single text.

In this paper, we have built our own corpus in Hinglish which helps us to classify the early depression. We have annotated and labeled the entire data, as well as POS tagged it. Our models were built on the review data set to make it more generalize in classifying the depression.

2 Problem Statement

The major problem in code-switching or multi-lingual text is the grammatical context concerning particular language and how they affect the context of other languages. Because of the change in grammatical context, emotions related to that particular text also changes with it. For instance in example in S1, ”band” might be a more of neutral word, but if we look at T1 then it can be negative sentence correlating negative emotions.But that’s not the end here, many times people speak in sarcasm which is hard to identify. Let’s take the same example S1: ”Mahi will get f***ed up now”, this can be taken as a sarcasm in Hindi as well as English.

Because of all these fundamental problems in Natural Language Processing, it becomes really hard to classify the depression in multilingual text. For instance depress people can talk or write more in their native language just because they feel more comfortable in it, on the other hand some people talk in mix language without even realising due to more depress and tense behaviour. This becomes more of psycho-linguistic problem. We know everyone behaves and act differently which affects their language. In this paper we have addressed this problem and came up with our psycho-linguistic solution.

3 Proposed Methodology

3.1 Data Collection and Cleaning

Using the twitter API and other scrapping tools (mentioned in appendix A), we extracted around 821 tweets with the following hashtags: #depression, #depressed, #happiness. Since our focus is on code-switching between Hindi and English, most of the tweets were posted by people of Indian origin. A particular date range was set while scraping. Table 1 shows the dataset statistics. Tweets with the hashtags #depressed and #depressionIsReal contribute to approximately 36% of the dataset while the remaining 54% have hashtags #happiness or #happy. This imbalance is because people are reluctant to discuss mental health issues on social media platforms.

3.1.1 Pre-processing steps

1. **Removal of URLs and usernames**: Regular expression worked well in this case.

Table 1: **Data Distribution**

Emotion	Sentences
Depressed	36.55%
Happy	63.45%

2. **Emoticons**: A mix of emoticons were used by the users. Some of them were from the set provided by twitter and others were similar to these: ':)', ':(', ':P', ':/'. While these could be one of the features that a statistical model learns from, the inconsistency caused some errors and we removed them for now.

3. **Romanized Hindi to Devnagri Transliteration**: The tweets obtained were written in mixed scripts i.e. some of the Hindi words are written in their Romanized equivalents. By identifying the language (as Hindi) of the word token, we used Google Translate API to get the Devnagari equivalent for all the tokens. Thus, we keep the uniform representation of Hindi and English by transliterating each in their respective scripts.

3.2 Corpus Creation and Annotation

3.2.1 Language Annotation

(Sreeram et al.,) has described various instances of code-switching and occurrence in sentences in their corpus. Since we have considered intra-sentence code-switching, each word needs to be annotated using language identification. NLTK's English corpus is a good resource in this step. Using this we obtained a dictionary of all the possible words in the English language and tagged each word that did not belong to this dictionary as 'unusual'. Once this was done, the list of unusual words contained a mix of Hindi words and some misspelled English ones. These misspelled words were filtered out. An example of language annotation is shown below.

"माहि":"hin", "का" : "hin", " तो" : "hin", " band" : "eng", " बज" : "hin", " जाएग : "hin"

3.2.2 Causal Language Annotation

While training on code-switched data one important feature is the language of emotion. Emotions can be expressed in the Hindi and English separately and also through mixed phrases such as "band baj jaega". We annotated each sentence as either "Hindi", "English" or "Mixed". **Hindi** means that the emotion in the sentence is solely expressed through Hindi. For example:

S: "आज फिर वही जानी पहचानी सी बेचैनी होने लगी"
T: "Today again the same familiar restlessness started happening"

Similarly, a sentence would be annotated as **English** if the emotion is solely expressed through English. For example:

S: "यूँ तो ज़िन्दगी frustrated, depressed बस चल रही है न चाहने पर भी उपरवाले के भी"
T: "Life is frustrating and depressing even if its against God's wishes"

From Table 2 it can be seen that, Hindi is dominating as the causal language in most of the tweets. This is because people prefer to switch to their native language while expressing their emotions as discussed in (Caldwell-Harris, 2014).

Table 2: **Caused Language Statistics**

Causal Language	Sentences
Hindi	228 (76%)
English	54 (18%)
Mixed	16 (5%)

3.3 Feature Extraction

3.3.1 Negation Words

Negation words could cause a variance from the actual emotion. For example, in the phrase "not happy", the negation word "not" is causing a shift from the emotion of happiness. So the sentence "I am not happy even though its the holiday season" could wrongly be classified as belonging to the "not depressed" class if the model only learns to look at keywords such as "happy".

We made two separate lists of negation

words for the two languages. The number of negation words in each sentence was counted and this count was given as a feature.

negationHin = 'मत', 'नहीं', 'न', 'ना', 'का'
negationEng = 'Doesn't', 'Isn't', 'Wasn't', 'Hardly', 'Scarcely', 'Barely', 'No', 'Not', 'Nowhere', 'Never', etc.

3.3.2 Probability of Being Positive

Another feature that we included was the probability of each word being positive. Two values were calculated:
1. the frequency of the word in the depression dataset **(a)** divided by the sum of its frequency in each of the datasets **(a+b)**.
2. the frequency of the word in the non-depression dataset **(b)** divided by the sum of its frequency in each of the datasets **(a+b)**.

The maximum of the above two was assigned as Prob(word) given by the following :

$$max \frac{a}{a+b}, \frac{b}{a+b} \tag{1}$$

3.3.3 Character N-Grams

As discussed in (Kulmizev et al., 2017), character n-grams work especially well in the case of code-switching because they are not language-dependent. Since this dataset has an informal tone and has been taken from a social media platform, there are cases where words have been misspelled. This issue can be handled by character n-grams pretty well.

3.3.4 Word N-Grams

3.4 Classification Methods

We have used different machine learning classification algorithms to predict the depression in Hindi-English mixed sentences. We trained the models on Hindi and English and Hinglish Review datasets as well as on the Hinglish depression dataset.

3.4.1 Logistic Regression

Logistic Regression approach estimates the probabilities using logistic function that models binary dependent variable and one or more independent variables. It is a type of generalized linear methods, only the predictions are transformed using the logistic or sigmoid function. Logistic regression is a statistical analysis method for analyzing the data where there are one or more than independent variables that determine the the class.

3.4.2 Naive Bayes

Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with naive independence assumptions between various features. All naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. An advantage of naive Bayes is that it only requires a small number of training data to estimate the parameters necessary for classification.

3.4.3 Support Vector Machine

Support Vector Machine(SVM) algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. It constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space. SVMs perform well in text and hypertext categorization, as their application can significantly reduce the need for labeled training instances in both the standard inductive and transductive settings.

3.4.4 Multilayer Perceptron

A multilayer perceptron(MLP) is a class of feedforward deep neural network. An MLP consists of at least three layers of nodes: an input layer, a hidden layer and a output layer. Each node in a hidden and output layers is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation while training. It can distinguish data that is not linearly separable. It is fully connected acyclic graph where each node in one layer connects with a certain weight w to every node in the following layer.

3.4.5 Decision Tree

Decision Tree(DT) is a flowchart-like structure in which each internal node represents a 'test' on an attribute and each branch represents the outcome of that test, and each leaf node represents a class label. The paths from root to leaf represent classification rules.

3.4.6 Ensemble Method

Ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. The trained ensemble represents a single hypothesis. Ensembles tend to yield better results when there is a significant diversity among the models. A necessary and sufficient condition for an ensemble of classifiers to be more accurate than any of its individual members is if the classifiers are accurate and diverse.

4 Experiments and Results

4.1 Transfer Learning

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. It is an approach to try if a related task with abundant data can be identified. This optimization technique allows rapid progress or improved performance when modeling the second task.

The depression dataset, is not large enough to train a good model. Thus, we took help from another related task that had available an abundant dataset called the Review Dataset. https://github.com/shubham721/Sentiment-Analysis-On-Hindi-Reviews.

Table 3: **Review Dataset**

Model	Accuracy	F-Score
Log Regression	0.9233	0.9601
SGD	0.9395	0.9699
SVC	0.9233	0.9601
Nearest Neighbour	0.9157	0.9558
MLP	0.9405	0.9692
DecisionTree	0.9286	0.9638
MultinomialNB	0.9233	0.9601
Ensemble Classifier	0.9416	0.9693

We observed that transfer learning helped us to improve the accuracy for our Hinglish sentences. As had limited Hinglish depression related sentences for training, transfer learning enabled us to acquire knowledge from similar sentiments and apply them to related depression data. Our model learned the emotional relationship of positive and negative related to other people.

Binary sentiment classification (BSC) on a code-switched review dataset is semantically similar to depression detection (also a binary sentiment classification task).

So formally the above statement can be expressed as follows:
Soucre task: BSC on code-switched review dataset
Target task: BSC on code-switched depression dataset

There are multiple methods to apply transfer learning. Here, we use the Parameter Initialization technique which trains on the source task and directly uses the tuned parameters to initialize the network for the target task. After the transfer, a fine-tuning is performed on the parameters.

Table 4: **Depression in Hinglish**

Unigram+Tfidf	Accuracy	F-Score
Logistic Reg	0.9076	0.9424
SGD	0.9117	0.936
SVC	0.8974	0.9427
Nearest Neighbour	0.8877	0.9223
MLP	0.9163	0.946
DecisionTree	0.8987	0.9509
MultinomialNB	0.9027	0.9335
Ensemble Classifier	0.9206	0.9573

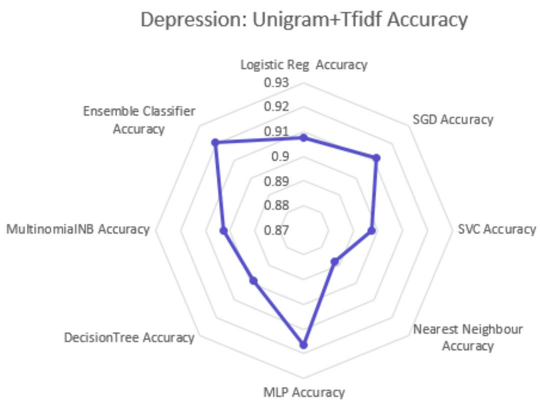

Figure 1: Radar plot for model-wise Accuracy on Depression Dataset

For the depression dataset, we got the high-

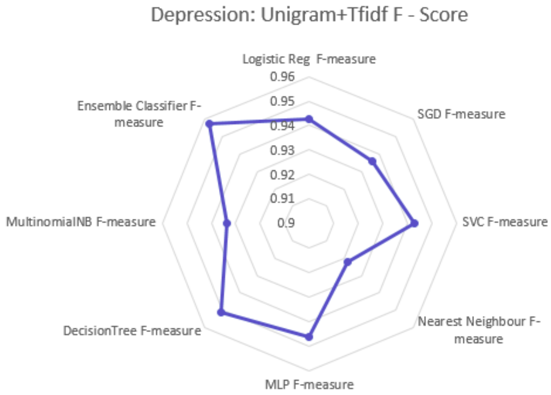

Figure 2: Radar plot for model-wise F-Scores on Depression Dataset

est accuracy of 92.27% with ensemble method. The multilayer perceptron also performed well for classification with 91.84% of accuracy. Refer the Figure 1 for the further details.

5 Related Work

Emotion Prediction is a Natural Language Processing task dealing with the detection and classification of emotions in various monolingual and bilingual texts. While some work has been done on code-mixed social media text and in emotion prediction separately, (Vijay et al., 2018) aims to identify the emotion associated with Hindi-English code-mixed social media text. they have analyzed the problem of emotion identification in code-mixed content and present a Hindi-English code-mixed corpus extracted from twitter and annotated with the associated emotion. Whereas authors of (Singh et al., 2018) presented a unique language tagged and POS-tagged dataset of code-mixed English-Hindi tweets related to five incidents in India that led to a lot of Twitter activity with a POS tagging model.

In our analysis and study, we have observed that code-switched documents have also received considerable attention in the NLP community. Several studies have focused on identification and analysis, including mining translations in code-switched documents, predicting code-switched points, identifying code-switched tokens, adding code-switched support to language models, and developing POS tagging for code-switching text. (Wang et al., 2017) have proposed a joint factor graph model that addresses the issue of different emotions presented in a text. They have proposed an attribute functions of the factor graph model to utilized and learn both monolingual and bilingual information from each text, factor functions were used to explore the relationship among different emotions, and a belief propagation algorithm was employed to learn and predict the model. Their empirical studies demonstrate the importance of emotion analysis in code-switching text and the effectiveness of our proposed joint learning model.

Whereas in (Conneau et al., 2019) they have shown that pre-training multilingual language models at scale lead to significant performance gains for a wide range of cross-lingual transfer tasks. The possibility of multilingual modeling without sacrificing per-language performance, XLM-R is very competitive with strong monolingual models on the GLUE and XNLI benchmarks. (Trotzek et al., 2018) has addressed the early detection of depression by using machine learning models based on messages on a social platform. In particular, a convolutional neural network based on different word embedding is evaluated and compared to a classification based on user-level linguistic metadata.

End-to-end (E2E) systems are fast replacing the conventional systems in the domain of automatic speech recognition. As the target labels are learned directly from speech data, the E2E systems need a bigger corpus for effective training. Authors of (Dhawan et al., 2019) have done this by reducing the number of target labels for reliable training of the E2E systems on limited data. The efficacy of the proposed approach has been demonstrated on two prominent architectures, namely CTC based and attention-based E2E networks. The experimental validations are performed on a recently created Hindi-English code-switching corpus.

6 Conclusion

In this paper, we present a freely available, small Hindi-English code-switched corpus for depression classification. It consists of cleaned code-mixed tweets with corresponding language annotations. We also present various supervised machine learning models which then used to classify the texts into *depressed/ happy* categories. We observed that, features such as character n-grams, word n-grams are important in identifying the context in code-switched corpus. One of the findings of this work is the use of transfer learning in which we used a pretrained model on different code-switched dataset and fine tuned the model for our application. Due to this, the transfer learning helped in identifying the initial level language features, such as structures, word and language usage. This helped us in improving our classification accuracy to 94.2%. This can be reiterated for many languages and the scope can be increased to consider more and more languages.

As a part of future work, the POS tagged corpus can be further enriched with grammar level features and by adding this to a larger corpus, it can be used to train automatic POS taggers in code-switched contexts. In a bigger context, we would like to build a system in which a psychiatrist uses this early depression detection system in live environment by incorporating facial emotion recognition system with Automatic speech recognition for tone detection.

Acknowledgments

We would like to acknowledge our professor Dr. John Hale for guiding in this project. Also, would like to thank Dr. Philip Resnik for motivating to explore the linguistic aspect of depression studies.

References

Catherine L. Caldwell-Harris. 2014. Emotionality differences between a native and foreign language: theoretical implications. *Frontiers in Psychology*, 5.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Kunal Dhawan, Ganji Sreeram, Kumar Priyadarshi, and Rohit Sinha. 2019. Investigating target set reduction for end-to-end speech recognition of hindi-english code-switching data. *arXiv preprint arXiv:1907.08293*.

Artur Kulmizev, Bo Blankers, Johannes Bjerva, Malvina Nissim, Gertjan van Noord, Barbara Plank, and Martijn Wieling. 2017. The power of character n-grams in native language identification. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 382–389, Copenhagen, Denmark, September. Association for Computational Linguistics.

Kushagra Singh, Indira Sen, and Ponnurangam Kumaraguru. 2018. A twitter corpus for hindi-english code mixed pos tagging. In *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*, pages 12–17.

Ganji Sreeram, Kunal Dhawan, and Rohit Sinha. *arXiv preprint arXiv:1810.00662*.

Marcel Trotzek, Sven Koitka, and Christoph M Friedrich. 2018. Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. *arXiv preprint arXiv:1804.07000*.

Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. Corpus creation and emotion prediction for hindi-english code-mixed social media text. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 128–135.

Zhongqing Wang, Sophia Yat Mei Lee, Shoushan Li, and Guodong Zhou. 2017. Emotion analysis in code-switching text with joint factor graph model. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 25(3):469–480.

A Supplemental Material

Following are the APIs we used:

1. `https://pypi.org/project/twitterscraper/0.2.7/`
2. `https://github.com/twintproject/twint.git`
3. `https://pypi.org/project/googletrans/`