Detecting Offensive Tweets in Hindi-English Code-Switched Language

Puneet Mathur
Netaji Subhas Institute of Technology
Delhi, India
pmathur3k6@gmail.com

Rajiv Ratn Shah
IIIT-Delhi
Delhi, India
rajivratn@iiitd.ac.in

Ramit Sawhney
Netaji Subhas Institute of Technology
Delhi, India
ramits.co@nsit.net.in

Debanjan Mahata
University of Arkansas at Little Rock
Arkansas, USA
dxmahata@ualr.edu

Abstract

The exponential rise of social media websites like Twitter, Facebook and Reddit in linguistically diverse geographical regions has led to hybridization of popular native languages with English in an effort to ease communication. The paper focuses on the classification of offensive tweets written in Hinglish language, which is a portmanteau of the Indic language Hindi with the Roman script. The paper introduces a novel tweet dataset, titled Hindi-English Offensive Tweet (HEOT) dataset, consisting of tweets in Hindi-English code switched language split into three classes: non-offensive, abusive and hate-speech. Further, we approach the problem of classification of the tweets in HEOT dataset using transfer learning wherein the proposed model employing Convolutional Neural Networks is pre-trained on tweets in English followed by retraining on Hinglish tweets.

1 Introduction

The rampant use of offensive content on social media is destructive to a progressive society as it tends to promote abuse, violence and chaos and severely impacts individuals at different levels. Offensive text can be broadly classified as abusive and hate speech on the basis of the context and target of the offense. Hate speech (Schmidt and Wiegand, 2017) is an act of offending, insulting or threatening a person or a group of similar people on the basis of religion, race, caste, sexual orientation, gender or belongingness to a specific stereotyped community. Abusive speech categorically differs from hate speech because of its casual motive to hurt using general slurs composed of demeaning words. Both abusive as well as hate speech are sub-categories of offensive speech.

Freedom of expression is one of the most aggressively contested rights of the modern world. While censorship of free moving online content such as Twitter tweets curtails the freedom of speech, but unregulated opprobrious tweets discourage free discussions in the virtual world (Silva et al., 2016). Hate speech detection is a hard research problem because of ambiguity in the clear demarcation of offensive, abusive and hateful textual content due to variations in the way people express themselves in a linguistically diverse social setting. A major challenge in monitoring online content produced on social media websites like Twitter, Facebook and Reddit is the humongous volume of data being generated at a fast pace from varying demographic, cultural, linguistic and religious communities.

A major contributor to the tremendously high offensive online content is *Hinglish* (Sreeram and Sinha, 2017), which is formed of the words spoken in Hindi language but written in Roman script instead of the Devanagari script. Hinglish is a pronunciation based bi-lingual language that has no fixed grammar rules.

Hinglish extends its grammatical setup from native Hindi accompanied by a plethora of slurs, slang and phonetic variations due to regional influence. Randomized spelling variations and mul-

Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media , pages 18–26,
Melbourne, Australia, July 20, 2018. ©2018 Association for Computational Linguistics

tiple possible interpretations of Hinglish words in different contextual situations make it extremely difficult to deal with automatic classification of this language. Another challenge worth consideration in dealing with Hinglish is the demographic divide between the users of Hinglish relative to total active users globally. This poses a serious limitation as the tweet data in Hinglish language is a small fraction of the large pool of tweets generated, necessitating the use of selective methods to process such tweets in an automated fashion. We aim to solve the problem of detecting offensive Hinglish tweets through the development of a deep learning model that analyses the input text and segregates them as:

1. Not Offensive
2. Abusive
3. Hate-Inducing

A dataset of manually annotated Hinglish tweets is used to measure the performance of the proposed framework. The experimentation consists of two phases, the first of which investigates the semantic correlation of Hindi-English code switched language with native English language and proposes a dictionary-based translation of Hinglish text into Roman English text. Next, we analyze the performance of the semantically similar but syntactically different tweets obtained via transliteration and translation on a pre-trained Convolutional Neural Network (CNN) and propose improvements to the classical hate speech classification methodology through the transfer of previously learned features by the CNN. The main contributions of our work can be summarized as follows:

- Creation of an annotated dataset of Hinglish tweets
- Experimentation of transfer learning based neural networks for classifying tweets in Hinglish language as abusive, hate-inducing or non-Offensive.

2 Related Work

The voluminous data present on Twitter necessitates identification, ranking and segregation of event-specific informative content from the streams of trending tweets (Mahata et al., 2015). Orsini (2015) dates the origin of Hinglish as an informal language to postcolonial Indian society. Several work like that done by Dwivedi and Sukhadeve (2010) attempted to translate Hindi-English language into pure English. However, the major challenge in this case is that the grammatical rules of Hinglish are gravely uncertain and user dependent.

One of the earliest efforts in hate speech detection can be attributed to Spertus (1997) who had presented a decision tree based text classifier for web pages with a remarkable 88.2 % accuracy. Contemporary works on Yahoo news pages were done Sood et al. (2012) and later taken up by Yin et al. (2016a) . Xiang et al. (2012) detected offensive tweets using logistic regression over a tweet dataset with the help of a dictionary of 339 offensive words. Offensive text classification in other online textual content have been tried previously for other languages as well like German (Ross et al., 2017) and Arabic (Mubarak et al., 2017). However, despite the various endeavors by language experts and online moderators, users continue to disguise their abuse through creative modifications that contribute to multidimensional linguistic variations (Clarke and Grieve, 2017).

Badjatiya et al. (2017) used CNN based classifiers to classify hateful tweets as racist and sexist. Park and Fung (2017) introduced a combination of CharCNN and WordCNN architectures for abusive text classification. Gambäck and Sikdar (2017) explored four CNN models trained on character n-grams, word vectors based on semantic information built using word2vec, randomly generated word vectors, and word vectors combined with character n-grams to develop a hate-speech text classification system. Mahata et al. (2018) experimented with multi-channel CNN, BiLSTM and CNN+BiLSTM models for identifying specific posts from a large dataset of Twitter posts. Another interesting attempt in the same direction was made by Pitsilis et al. (2018) through an ensemble model of Recurrent Neural Network (RNN) classifiers.

3 Dataset

Table 1 shows the tweet distribution in English dataset A provided byDavidson et al. (2017) and the manually created Hinglish dataset HEOT. Dataset A consists of 14509 tweets such that 7274 are non-offensive, 4836 are abusive and 2399 are hate-inducing tweets. The imbalance of dataset is

Label	Dataset A	Dataset HEOT
Non-Offensive	7274	1414
Abusive	4836	1942
Hate-inducing	2399	323
Total	14509	3679

Table 1: Tweet distribution in dataset A and HEOT.

encouraged to represent a realistic picture usually seen on social media websites.

Dataset HEOT was created using the Twitter Streaming API by selecting tweets in Hindi-English code switched language by data mining specific profane words in Hinglish language. The tweets were collected during the months of November-December 2017 and were crowd-sourced to ten NLP researchers for annotation and verification. The data repository thus created consists of 3679 tweets out of which the count of non-offensive, abusive and hate-inducing tweets is 1414, 1942 and 323 respectively and categorized similar to the previous dataset. Dataset HEOT is considerably small as compared to dataset A, but this abnormality is rather advantageous for our research. It is a common observation that online users who identify to a particular demographic subdivision are often a small percentage of the total active users. This restriction of the size of Hinglish corpus closely represents a true world scenario where the relative balance of standard and indigenous users is naturally skewed. Care was taken to ensure that the tweets having insufficient textual content were not incorporated into the dataset.

An illustration of the three types of tweets is presented below to explain the contextual meaning of each class label in different languages. The tweets in category 1, 2 and 3 are non-offensive, abusive and hate-inducing respectively. In the examples given here, each tweet belonging to class A and B is in English and Hinglish language respectively. The tweets that fall under class C exemplify the corresponding version of Hinglish tweets after transliteration, translation and preprocessing.

1. (a) We all are going outside? http://t...
   (b) Hum sab ghumne jaa rahe hain? http://t...
   (c) we all outside go are
2. (a) @username1 B*tch! Do not teach me:/
   (b) @username1 Kutiya! Mujhe mat sikha:/
   (c) b*tch me not teach
3. (a) M*th*rf*ck*r Kill terrorist Akbaar #SaveWorld
   (b) M*d*rch*d aatanki Akbaar ko maara daalo #SaveWorld
   (c) m*th*rf*ck*r terrorist Akbaar kill

Hinglish to Devanagari Hindi transliteration was done by using the datasets provided by Khapra et al. (2014), while the Hindi to Roman English translation was achieved by using the Hindi-English dictionary sourced from CFILT, IIT Bombay[1]. A crowdsourced list of 208 profane Hinglish words along with their spelling variations, regional dialects, homonyms and contextual variants were added to the corpus of 7193 word-pairs to be used for all the Hinglish to English tweet conversions discussed in this paper.

4 Methodology

4.1 Preprocessing

The tweets obtained from data sources were channeled through a pre-processing pipeline with the ultimate aim to transform them into semantic feature vectors.

The transliteration process was broken into intermediate steps:

1. Removal of punctuations, URLs and user mentions.

2. Replacement of hashtags with corresponding plain text.

3. Replacement of emoticons with appropriate textual descriptions sourced from the list provided by Agarwal et al. (2011).

4. Conversion of all tweets into lower case.

5. Removal of useless words providing little textual information using stop words obtained from Gensim (Rehurek and Sojka, 2011).

6. Translation of Hinglish words into corresponding English words.

[1] http://www.cfilt.iitb.ac.in/~hdict/webinterface_user/

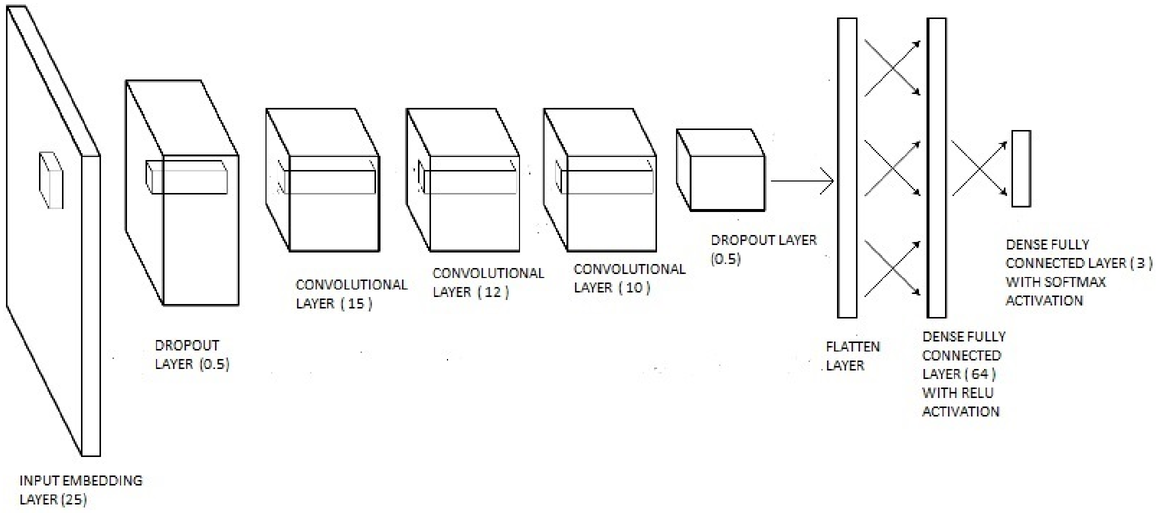
INPUT EMBEDDING LAYER (25)
DROPOUT LAYER (0.5)
CONVOLUTIONAL LAYER (15)
CONVOLUTIONAL LAYER (12)
CONVOLUTIONAL LAYER (10)
DROPOUT LAYER (0.5)
FLATTEN LAYER
DENSE FULLY CONNECTED LAYER (64) WITH RELU ACTIVATION
DENSE FULLY CONNECTED LAYER (3) WITH SOFTMAX ACTIVATION

Figure 1: Convolutional Neural Network (CNN) architecture used for Ternary Trans-CNN model

7. Transformation of pre-processed tweets into a word vector representation through Glove (Pennington et al., 2014) pre-trained vector embeddings. The version of Glove pre trained word vectors used in our case was Twitter (2B tweets, 27B tokens, 1.2M vocab, uncased, 200d, 1.42 GB download).

8. The final step in tweet transformation was the creation a word vector sequences that can be fed into the neural network architecture.

4.2 Transfer Learning

Transfer learning (Pan and Yang, 2010) is a machine learning paradigm that refers to knowledge transfer from one domain of interest to another, with the aim to reuse already learned features in learning a specialized task. The task from which the system extracts knowledge is referred to as source task while the task which benefits is termed as target task. Such representation learning systems are used in cases where the feature space and distribution of input are similar so as to get maximum benefit from the knowledge transfer exercise. Another pertinent role of transfer learning is data reclassification without overfitting in cases where data extraction restraints the size of training data.

Bengio (2012) put emphasis on two predominant cases which are well suited for the application of transfer learning. The first case is when the class labels of source and target task vary but the input distribution is same. The other is when the class labels are similar but the input distribution varies. The proposed problem of hate speech detection in Hinglish tweets is a classic example of the second case due to the semantic parallelism between English and translated Hinglish language, despite the eventual grammatical disassociation when Hinglish is transliterated into Roman script. Transfer learning provides relative performance increase at a reduced storage and computational cost.

Pan and Yang (2010) gave a mathematical definition of transfer learning and justified use cases for application of transfer learning. Let domain D consist of two components: a feature space X and a marginal probability distribution $P(X)$, where $X = \{x_1, x_2, \cdots, x_n\} \in X$, X is the space of all individual word vectors representing the input text, x_i is the i^{th} vector corresponding to some tweet and X is a particular learning sample. A task consists of two components: a label space Y and an objective predictive function $f(\dot{})$, represented as $T = \{Y, f(\dot{})\}$, which is not observed but can be learned from the training data, which consists of pairs (x_i, y_i), where $x_i \in X$ and $y_i \in Y$. In the experiments, Y is the set of all labels for a multi-class classification task, and y_i is one of three class labels. Figure 1 shows the architecture of convolutional neural network used in the experiments throughout the paper. CNN models pre-trained on English dataset learn low-level features of the English language. The last few layers are removed and then replaced with fresh layers keeping the initial convolutional layers frozen and retrained on dataset HEOT where it learns to ex-

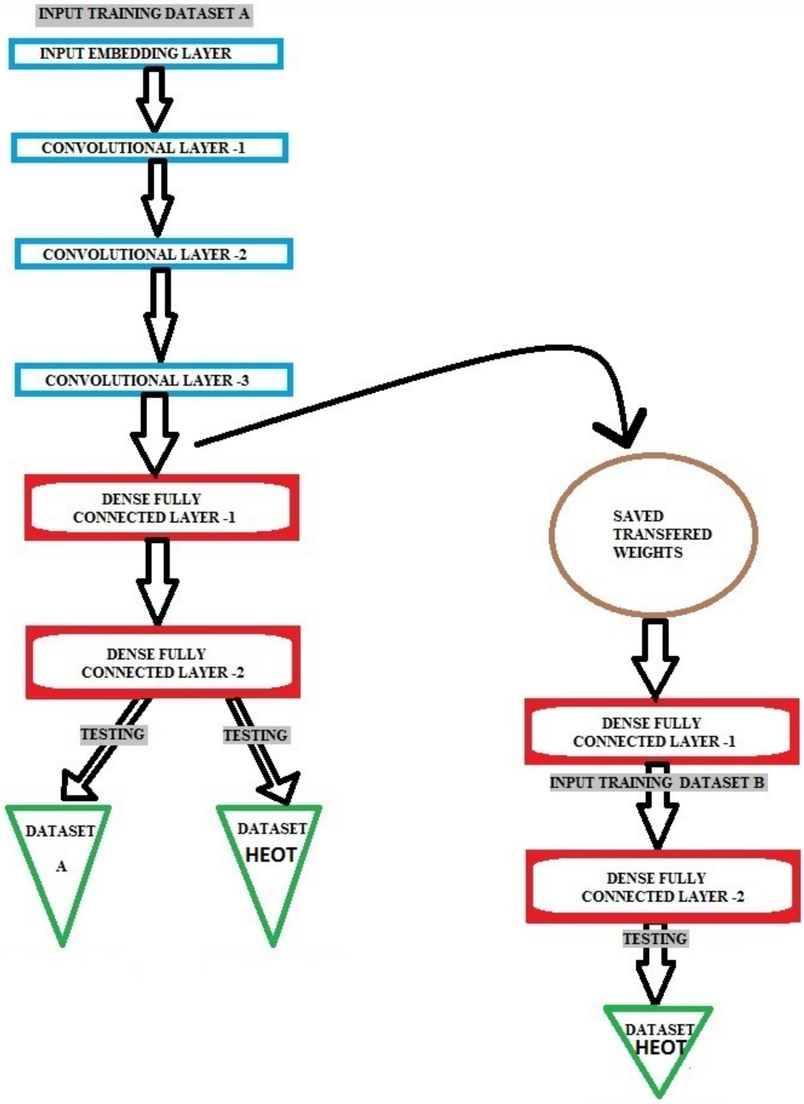

Figure 2: Transfer learning technique used for Ternary Trans-CNN model

tract intricate features due to syntax variations in pre-processed Hinglish text.

5 Proposed Approach

The authors have put forward an experimental schema for Hinglish hate speech classification, termed as Ternary Trans-CNN model.

5.1 Ternary Trans-CNN Model

Ternary Trans-CNN model aims to achieve the three-label classification of Hinglish tweets using transfer learning on a pre-trained CNN architecture depicted in Figure 2. The model is trained successively on English dataset A and Hinglish dataset HEOT. We have empirically chosen embedding dimension to be 200. The proposed CNN architecture consists of 3 layers of Convolutional1D layers having filter size 15,12 and 10 respectively and kernel size fixed to 3. The last two layers are dense fully-connected layers with size 64 and 3 units and the activation function as Rectified Linear Unit (ReLU) (Maas et al., 2013) and 'Softmax' respectively. The loss function used is categorical cross-entropy on account of multi-label classification role of the model. We used Adam optimizer (Kingma and Ba, 2014)

The batch size was experimented from size 8 to 256 using grid search. Similarly, the number of epochs were chosen by exploring different values from 10 to 50. The number of trainable and static layers varied to get the best combination giving optimal results. To ensure that the models do not overfit, dropout layers after the dense layers were

Dataset	A	HEOT (w/o TFL)	HEOT (TFL)
Accuracy (%)	75.40	58.70	83.90
Precision	0.672	0.556	0.802
Recall	0.644	0.473	0.698
F1 Score	0.643	0.427	0.714

Table 2: Results for Ternary Trans-CNN task: non-offensive, abusive and hate-inducing tweet classification on datasets A, HEOT without transfer learning (w/o TFL) and HEOT with transfer learning (TFL)

introduced to enhance generalization of the systems. The Ternary Trans-CNN model is initially trained on 11509 training data points and tested on 3000 data points that were randomly split from the parent dataset A. The batch size is set to 128 for 25 epochs with all layers as trainable. The same model is retrained, keeping only last two layers as trainable and other layers frozen, on dataset HEOT which is split into 2679 training and 1000 testing examples. The batch size was decreased to 64 with epochs reduced to 10 for minimum training loss and the metric measurements were recorded in Table 2 for further comparative analysis.

6 RESULTS AND ANALYSIS

The results of Ternary Trans-CNN model were compiled in terms of accuracy, F1 score, precision, and recall by choosing macro metrics as the class imbalance is not severe enough to strongly bias the outcomes. The CNN model was initially trained on dataset A and its performance on it taken as the baseline. Testing the same model on dataset HEOT without transfer learning reports downfall in model performance as compared to the baseline which is justified because the Hinglish tweets in dataset HEOT suffer from syntactic degradation after transliteration and translation which leads to a loss in the contextual structuring of the tweets. After retraining the Trans-CNN model, the model performance on dataset HEOT not only improves significantly but also surpasses the earlier results on dataset A. Thus, we can safely conclude that there was a positive transfer of features from source to target data.

7 CONCLUSION AND FUTURE WORK

This work demonstrates various CNN based models for multi-class labeling of offensive textual tweets. An important contribution of the paper is to analyze informal languages on social media such as Hinglish for hate speech and suggest ways to transform them into English text for the purpose of natural language processing. The dataset provided is an optimistic step in contribution to the study of code-switched languages such as Hinglish that play a major role in online social structuring of multi-linguistic societies. The experiments prove that a positive transfer of knowledge and characteristics between two congruent domains is made possible by training, freezing and retraining the models from source to target tasks. The success of transfer learning for analyzing complex cross linguistic textual structures can be extended to include many more tasks involving code-switched and code-mixed data.

The future efforts can be directed towards fine-tuning the neural network models using boosting methods such as gradient boosting (Badjatiya et al., 2017). The experiments here used CNN models for primary training, but other types of deep learning models like LSTM have also been known to show a high affinity for semantic tasks such as sentiment analysis (Wang et al., 2016) and sentence translation (Sutskever et al., 2014). Another possible approach to fine-tune the classification can be to use a stacked ensemble of shallow convolutional neural network (CNN) models as shown by Friedrichs et al. (2018).

In recent years, leveraging multimodal information in several multimedia analytics problem has shown great success (Shah, 2016a,b; Shah and Zimmermann, 2017). Thus, in the future, we plan to exploit multimodal information in offensive language detection since the most of existing systems work in unimodal settings. Moreover, since offensive language is closely related with sentiments, keywords (or hashtags), and some associated events, we would also like to explore aspects (Jangid et al., 2018), tag relevance (Shah et al., 2016a,b), and events (Shah et al., 2015a, 2016c) for the present problem. Furthermore, we would like extend our work to build an offensive video segmentation system (Shah et al., 2014a, 2015b) in order to filter abusive and hate-inciting videos on social media. Since offensive code-switched languages are heavily influenced by region (*i.e.*, location), we would try to exploit the location information of videos as well in our extended work (Shah et al., 2014b,c; Yin et al., 2016b). Finally, since relative positions of words play a pivotal role in analyzing Hindi, we would like explore such possibilities in our future work (Shaikh et al., 2013a,b).

References

Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In *Proceedings of the workshop on languages in social media*, pages 30–38. Association for Computational Linguistics.

Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, and Vasudeva Varma. 2017. Deep learning for hate speech detection in tweets. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 759–760. International World Wide Web Conferences Steering Committee.

Yoshua Bengio. 2012. Deep learning of representations for unsupervised and transfer learning. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, pages 17–36.

Isobelle Clarke and Jack Grieve. 2017. Dimensions of abusive language on twitter. In *Proceedings of the First Workshop on Abusive Language Online*, pages 1–10.

Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. *arXiv preprint arXiv:1703.04009*.

Sanjay K Dwivedi and Pramod P Sukhadeve. 2010. Machine translation system in indian perspectives. *Journal of computer science*, 6(10):1111.

Jasper Friedrichs, Debanjan Mahata, and Shubham Gupta. 2018. Infynlp at smm4h task 2: Stacked ensemble of shallow convolutional neural networks for identifying personal medication intake from twitter. *arXiv preprint arXiv:1803.07718*.

Björn Gambäck and Utpal Kumar Sikdar. 2017. Using convolutional neural networks to classify hate-speech. In *Proceedings of the First Workshop on Abusive Language Online*, pages 85–90.

Hitkul Jangid, Shivangi Singhal, Rajiv Ratn Shah, and Roger Zimmermann. 2018. Aspect-based financial sentiment analysis using deep learning. In *Companion of the The Web Conference 2018 on The Web Conference 2018*, pages 1961–1966. International World Wide Web Conferences Steering Committee.

Mitesh M Khapra, Ananthakrishnan Ramanathan, Anoop Kunchukuttan, Karthik Visweswariah, and Pushpak Bhattacharyya. 2014. When transliteration met crowdsourcing: An empirical study of transliteration via crowdsourcing using efficient, non-redundant and fair quality control. In *LREC*, pages 196–202.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. 2013. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3.

Debanjan Mahata, Jasper Friedrichs, Rajiv Ratn Shah, et al. 2018. # phramacovigilance-exploring deep learning techniques for identifying mentions of medication intake from twitter. *arXiv preprint arXiv:1805.06375*.

Debanjan Mahata, John R Talburt, and Vivek Kumar Singh. 2015. From chirps to whistles: discovering event-specific informative content from twitter. In *Proceedings of the ACM web science conference*, page 17. ACM.

Hamdy Mubarak, Kareem Darwish, and Walid Magdy. 2017. Abusive language detection on arabic social media. In *Proceedings of the First Workshop on Abusive Language Online*, pages 52–56.

Francesca Orsini. 2015. Dil maange more: Cultural contexts of hinglish in contemporary india. *African Studies*, 74(2):199–220.

Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

Ji Ho Park and Pascale Fung. 2017. One-step and two-step classification for abusive language detection on twitter. *arXiv preprint arXiv:1706.01206*.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Georgios K Pitsilis, Heri Ramampiaro, and Helge Langseth. 2018. Detecting offensive language in tweets using deep learning. *arXiv preprint arXiv:1801.04433*.

R Rehurek and P Sojka. 2011. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2).

Björn Ross, Michael Rist, Guillermo Carbonell, Benjamin Cabrera, Nils Kurowsky, and Michael Wojatzki. 2017. Measuring the reliability of hate speech annotations: The case of the european refugee crisis. *arXiv preprint arXiv:1701.08118*.

Anna Schmidt and Michael Wiegand. 2017. A survey on hate speech detection using natural language processing. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, pages 1–10.

Rajiv Ratn Shah. 2016a. Multimodal analysis of user-generated content in support of social media applications. In *Proceedings of the ACM International Conference on Multimedia Retrieval*, pages 423–426.

Rajiv Ratn Shah. 2016b. Multimodal-based multimedia analysis, retrieval, and services in support of social media applications. In *Proceedings of the ACM International Conference on Multimedia*, pages 1425–1429.

Rajiv Ratn Shah, Anupam Samanta, Deepak Gupta, Yi Yu, Suhua Tang, and Roger Zimmermann. 2016a. PROMPT: Personalized user tag recommendation for social media photos leveraging multimodal information. In *Proceedings of the ACM International Conference on Multimedia*, pages 486–492.

Rajiv Ratn Shah, Anwar Dilawar Shaikh, Yi Yu, Wenjing Geng, Roger Zimmermann, and Gangshan Wu. 2015a. EventBuilder: Real-time multimedia event summarization by visualizing social media. In *Proceedings of the ACM International Conference on Multimedia*, pages 185–188.

Rajiv Ratn Shah, Yi Yu, Anwar Dilawar Shaikh, Suhua Tang, and Roger Zimmermann. 2014a. ATLAS: Automatic temporal segmentation and annotation of lecture videos based on modelling transition time. In *Proceedings of the ACM International Conference on Multimedia*, pages 209–212.

Rajiv Ratn Shah, Yi Yu, Anwar Dilawar Shaikh, and Roger Zimmermann. 2015b. TRACE: A linguistic-based approach for automatic lecture video segmentation leveraging Wikipedia texts. In *Proceedings of the IEEE International Symposium on Multimedia*, pages 217–220.

Rajiv Ratn Shah, Yi Yu, Suhua Tang, Shin'ichi Satoh, Akshay Verma, and Roger Zimmermann. 2016b. Concept-level multimodal ranking of Flickr photo tags via recall based weighting. In *Proceedings of the MMCommon's Workshop at ACM International Conference on Multimedia*, pages 19–26.

Rajiv Ratn Shah, Yi Yu, Akshay Verma, Suhua Tang, Anwar Shaikh, and Roger Zimmermann. 2016c. Leveraging multimodal information for event summarization and concept-level sentiment analysis. *Knowledge-Based Systems*, 108:102–109.

Rajiv Ratn Shah, Yi Yu, and Roger Zimmermann. 2014b. ADVISOR: Personalized video soundtrack recommendation by late fusion with heuristic rankings. In *Proceedings of the ACM International Conference on Multimedia*, pages 607–616.

Rajiv Ratn Shah, Yi Yu, and Roger Zimmermann. 2014c. User preference-aware music video generation based on modeling scene moods. In *Proceedings of the ACM International Conference on Multimedia Systems*, pages 156–159.

Rajiv Ratn Shah and Roger Zimmermann. 2017. *Multimodal Analysis of User-Generated Multimedia Content*. Springer.

Anwar D Shaikh, Mukul Jain, Mukul Rawat, Rajiv Ratn Shah, and Manoj Kumar. 2013a. Improving accuracy of SMS based FAQ retrieval system. In *Proceedings of the Springer Multilingual Information Access in South Asian Languages*, pages 142–156.

Anwar Dilawar Shaikh, Rajiv Ratn Shah, and Rahis Shaikh. 2013b. SMS based FAQ retrieval for Hindi, English and Malayalam. In *Proceedings of the ACM Forum on Information Retrieval Evaluation*, page 9.

Leandro Araújo Silva, Mainack Mondal, Denzil Correa, Fabrício Benevenuto, and Ingmar Weber. 2016. Analyzing the targets of hate in online social media. In *ICWSM*, pages 687–690.

Sara Owsley Sood, Elizabeth F Churchill, and Judd Antin. 2012. Automatic identification of personal insults on social news sites. *Journal of the Association for Information Science and Technology*, 63(2):270–285.

Ellen Spertus. 1997. Smokey: Automatic recognition of hostile messages. In *AAAI/IAAI*, pages 1058–1065.

Ganji Sreeram and Rohit Sinha. 2017. Language modeling for code-switched data: Challenges and approaches. *arXiv preprint arXiv:1711.03541*.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Jin Wang, Liang-Chih Yu, K Robert Lai, and Xuejie Zhang. 2016. Dimensional sentiment analysis using a regional cnn-lstm model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 225–230.

Guang Xiang, Bin Fan, Ling Wang, Jason Hong, and Carolyn Rose. 2012. Detecting offensive tweets via topical feature discovery over a large scale twitter corpus. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1980–1984. ACM.

Dawei Yin, Yuening Hu, Jiliang Tang, Tim Daly, Mianwei Zhou, Hua Ouyang, Jianhui Chen, Changsung Kang, Hongbo Deng, Chikashi Nobata, et al. 2016a. Ranking relevance in yahoo search. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 323–332. ACM.

Yifang Yin, Rajiv Ratn Shah, and Roger Zimmermann. 2016b. A general feature-based map matching framework with trajectory simplification. In *Proceedings of the 7th ACM SIGSPATIAL International Workshop on GeoStreaming*, page 7.