

Identifying and Categorizing Opinions Expressed in Bangla Sentences using Deep Learning Technique

Moqsadur Rahman
Dept. of CSE
Shahjalal University of Science
and Tech., Sylhet

Summit Haque
Dept. of CSE
Shahjalal University of Science
and Tech., Sylhet

Zillur Rahman Saurav
Dept. of CSE
Shahjalal University of Science
and Tech., Sylhet

ABSTRACT

Identifying and categorizing opinions in a sentence is the most prominent branch of natural language processing. It deals with the text classification to determine the intention of the author of the text. The intention can be for the presentation of happiness, sadness, patriotism, disgust, advice, etc. Most of the research work on opinion or sentiment analysis is in the English language. Bengali corpus is increasing day by day. A large number of online News portals publish their articles in Bengali language and a few News portals have the comment section that allows expressing the opinion of people. Here a research work has been done on Bengali Sports news comments published in different newspapers to train a deep learning model that will be able to categorize a comment according to its sentiment. Comments are collected and separated based on immanent sentiment. The deep learning algorithms that have been used are Convolutional Neural Network (CNN), Multilayer Perceptron, Long Short-Term Memory (LSTM).

General Terms

Sentiment Analysis, Deep Learning, Emotion Classification

Keywords

CNN, LSTM, ROC curve, Confusion Matrix, Performance Analysis

1. INTRODUCTION

Sentiment analysis is the process of determining the emotional biases behind a series of words, used to achieve an understanding of the attitudes, opinions, and emotions expressed within an online mention. In general, sentiment analysis aims to determine the attitude of a speaker, writer or another subject concerning some topic or the overall contextual polarity or emotional reaction to a document, interaction or event. Sentiments are inherently subjective. Different people may interpret the attitude of the same text differently. It is extremely useful in social media monitoring as it allows us to find an overview of the wider public opinion behind certain topics.

The Internet is the global platform to share opinions. Thousands of people share their opinion on social media and blogs, post reviews on different products and services. The online news portal publishes its articles in different categories. In such a way, online text content is growing rapidly.

People need to decide by analyzing product reviews, news articles and social media posts and so on. It is also important to know individual journalists and columnist's opinions and public opinion on an issue like sports, business and trade, international, national and some other important issues. It is hard to analyze the opinions and feelings of huge news content manually. Sentiment Analysis focuses on determining

whether a piece of text is actually. It is also known as opinion mining, extracting the opinion or feelings of a writer. It is a process to detect someone's attitude towards a particular product or service.

Sentiment analysis is important to make decisions on topics like politics, sports, financial condition, product reviews and so on judgmental issues. Humans are subjective and that's why opinions are important. Business organizations, Government, Service-providing organizations, Sports lover analyze public opinions to identify opportunities, make a decision and make progress. The buyer wants to read reviews when to buy a product, Sports organization need to know public opinion to know the expectation of the audience, Government needs to know public opinion before making policy and even psychological investigation requires sentiment information. So, decision-makers are being depended on using the content of online media like-news, reviews, micro-blogs, and postings on social sites.

A newspaper shows us the current happenings of the world and can be called a mirror of society. The sports section of news portals shows news about Football, Cricket, Tennis, Hockey, sports club, sports team leaders and other players, score and sports issues. People read sports news to know the activity of sports scores, player's activities, sports organization policy, aim to achieve the goal in the future, local or national or international matters that will affect them. Some online news portals have comments sections where people can express their opinion that is helpful to know about other's opinion on a specific topic. Sentiment analysis can find the polarity of public opinion towards any topic. This study may help people to find current the condition of sports issues of any country and time.

People have different opinions in different situations. So, their expressions are different depending on the news they like or dislike. In Bengali news, people may share their comments as they like it most, some don't like it, some become worried and sad about it, some like to advise on it, some become angry and so on. So, there may be a huge amount of opinions expressed by people in Bengali news. In this paper, all those opinions are categorized by collecting those comments as data and labeling them as the people want to express. There are so many opinions or sentiment in Bengali comments like happiness (প্রসন্নতা), sadness (বিষন্নতা), desire (ইচ্ছা), compliment (অভিনন্দন), advice (উপদেশ), patriotism (দেশপ্রেম), annoyance (বিরক্তি), etc. From those sentiments, only four chosen for this work. They are happiness (প্রসন্নতা), sadness (বিষন্নতা), advice (উপদেশ) and annoyance (বিরক্তি). When some comment indicates any of the above four it means that comment or sentence have an opinion expressed by the people. When no opinions are implied from the above four,

we like to label them as neutral (নিরপেক্ষ). So comments or data are labeled using five classes after collected.

Various statistical and linguistic techniques have been developed for Sentiment analysis. All these methods are applied to the English language and there is a huge scope to work with the Bengali language. In this investigation, Neural Network models are used to detect the sentiment of Bengali sports news comments.

2. RELATED PREVIOUS WORKS

Identifying and categorizing opinions expressed in Bangla sentences is one of the common text classification problems and researches have been conducted on sentiment analysis on Bengali text.

Das & Bandyopadhyay presented two different approaches for identifying emotion holders from Bengali sentences. In this work, the first approach, the baseline model, is developed based on the combinations of various part-of-speech (POS) features extracted from the phrase-based similarities and the second approach, syntactic model, is based on the argument structure of the sentences concerning the verbs. They have tested Bangla text with cosine similarities using TF-IDF, Naive Bayes with POS tagger, stemmer. Some of them have worked on news article sentiment analysis [1].

The Centre for Research on Bangla Language Processing of BRAC University has conducted some works. They proposed a way to implement a corpus by collecting data from online resources [2]. They implemented a POS tagger based on HMM, n-gram and Brill's tagger [3]. The result is analyzed with a small corpus of 5000 words giving an accuracy of only 55%.

Md. Zahurul Islam and Naushad UzZaman present the compilation methodology and some statistical by observing the typical behavior of Zipf's curve for Bangla news corpus- "Prothom Alo corpus", which is the first of its kind for Bangla [4]. Muhammad Mahmudun Nabi, Tanzir Altaf and Sabir Ismail (2016) used feature sets and supervised classifiers and proposed a method to recognize the sentiment or opinion and extract a unique feature to come out a better approach to understanding sentiment from Bangla text using [5].

Horoscopes consist of future predictions for each of the twelve zodiac signs and are very popular in India. Tirthankar Ghosal and Sajal K. Das mainly focus on the sentiment analysis of Bengali daily horoscope using SVM with unigram features on the paper. They are given the positive and negative emotional basis of the sentence by crawling a leading Bengali newspaper's daily horoscope section [6].

MaxEnt and SVM algorithms are compared in paper [7] for sentiment analysis on Bengali microblog posts with different feature extraction methods and it gets the best performance with SVM with unigram and emoticons as features. Also, sentiment analysis is done on the Bengali horoscope corpus in paper [8] using ML algorithms. NB, SVM, K-Nearest Neighbors (NN), Decision Trees (DT), and Random Forest (RF). Among those SVM has the best performance with unigram features.

In paper [9] M. Trivedi, N. Soni, S. Sharma, and S. Nair shows how Support Vector Machine (SVM) and Naive Bayes (NB) algorithms are compared for text classification where SVM outperforms. S. Z. Mishu and S. M. Rafiuddin's Paper [10] compare text classification performance on different supervised ML models where back-propagation based Neural Network has the best performance.

3. DATASET

To start a Sentiment Analysis process, it is always required to build a sentiment lexicon and annotated data for machine learning. The main source of raw data of the corpus is web content. Online portals of more than 24 Bengali newspapers¹ are investigated for raw data and it has been found that Prothom Alo² has a huge collection of visitor's comments. So raw data of the corpus is collected from the popular Bengali newspaper Prothom-Alo.

After collecting all the raw data some data annotation campaigns were arranged where some students from the age range 19 to 25 annotated the data in five categories: happiness (প্রসন্নতা), sadness (বিষন্নতা), advice (উপদেশ), annoyance (বিরক্ত) and neutral (নিরপেক্ষ).

Here in the table 1 a brief statistics about the dataset is given,

Table 1. Dataset in brief

Source	www.prothomalo.com
Size	2492 sentences
Average word per sentence	8.3
Max word in a sentence	25

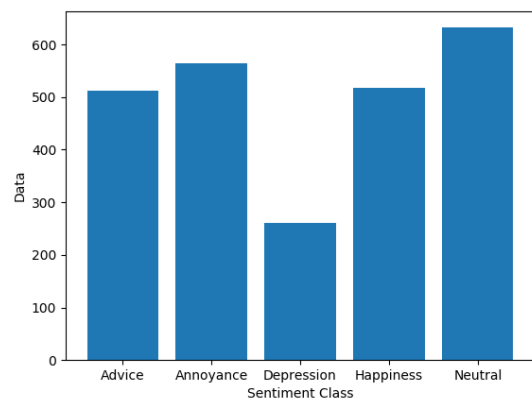


Figure 1. Number of data per class

In Figure. 1 we can get a brief overview of the number of data for each of the target classes. We can see that the dataset is quite imbalanced as the number of data in the "Depression" section is below 300 where all other classes have more than 500 data and the class labeled "Neutral" has more than 600 data. But since the dataset is human-annotated and all the labeling is done according to the sentiment of those sentences so nothing is done to maintain the balance.

4. DEEP NEURAL NETWORKS

Some traditional deep neural networks have been used in this experiment.

4.1 CNN

For CNN's strong ability to learn relevant features, its usage not only confined to image recognition, now it's heavily being used in NLP tasks. Different types of CNN have been used in this experiment. Here 1-D CNN, 1-D parallel CNN layers

¹ <http://www.24livenewspaper.com/bangla-newspaper>

² <http://www.prothom-alo.com>

with bi-gram, tri-gram, tetra-gram filters are used. A demo example of 1-D CNN with parallel layers where the embedding dimension is six is given in Figure 2. But in the experiments, different embedding dimensions have been tried and the best result is found using embedding dimension 200.

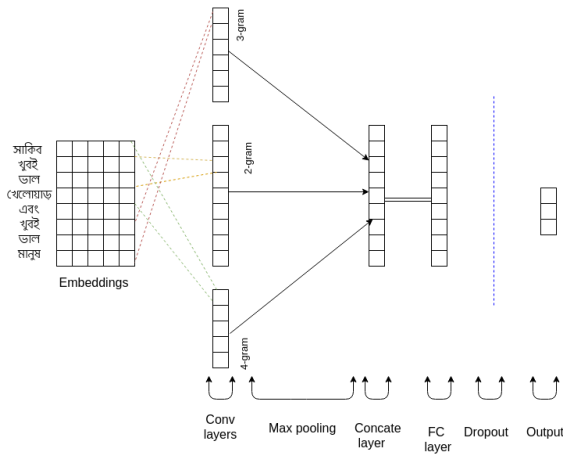


Figure 2. CNN with parallel layers

4.2 LSTM

RNNs are very efficient in capturing long-distance dependencies, and one of its variant LSTM solved its vanishing gradient problem [11]-[13]. LSTM cells carry past information through themselves. Each cell has different gates to interact with the data passing through it. LSTMs cell can update, remove the portion of data using the gates. A schematic design of a basic LSTM cell is given in Figure 3.

The equations used by an LSTM cell to perform operations are given below.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

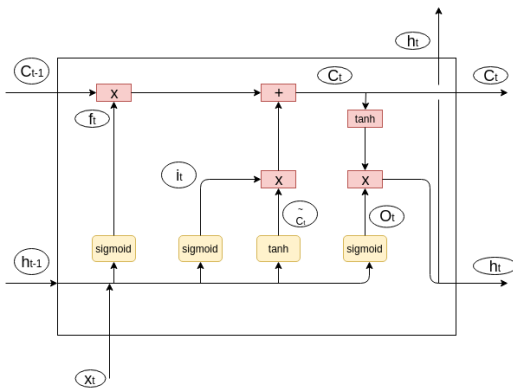


Figure 3. Basic LSTM cell

5. RESULTS & ANALYSIS

Deep learning models are applied to the dataset. Scikit-learn³ train-test split API has been used for splitting the dataset into train and test set and the ratio was 80 to 20. As it is preferred to use ROC curves measures for the imbalanced dataset, authors have principally used the F1 measures, ROC area measures for the datasets. ROC curve has a false positive rate on the X-axis and a True positive rate on the Y-axis. The rates were calculated using equation 7 and equation 8.

$$TPR = TP / (TP + FN) \quad (7)$$

$$FPR = FP / (FP + TN) \quad (8)$$

Here TPR, FPR, TP, FN, FP, TN stand for True Positive Rate, False Positive Rate, True Positive, False Negative, False Positive, True Negative respectively.

Experimenting with deep learning models on the dataset, some significant calculations are performed that are used as matrix for indicating the performance of those models. In Table 2 we will find F1 score, precision, and recall of all the three models.

Table 2. Performance measurement of all models

Model	F1 score	Precision	Recall
CNN	0.4819	0.4863	0.4819
LSTM	0.4658	0.4617	0.4658
DNN	0.4236	0.4319	0.4236

From the performance table, we can see that in every section the CNN model beats the other two models. This can be explained in the way that, LSTM works well in the scenario where sentences are long enough. In this dataset, the average word length is 8. So CNN worked here better than LSTM. So further analysis was continued on the CNN model. The ROC curve generated for the CNN model is shown in Figure 4.

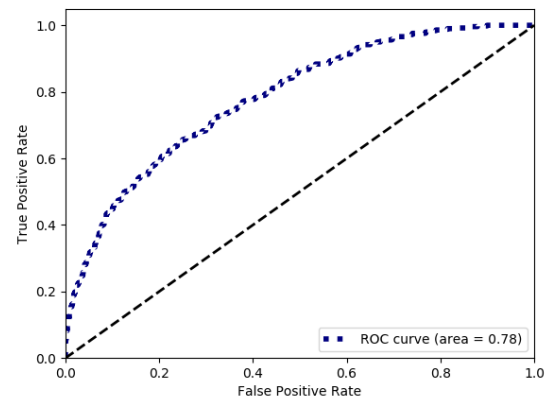


Figure 4. ROC curve for CNN

Confusion matrix for CNN model is shown in Figure 5.

³ <http://scikit-learn.org/>

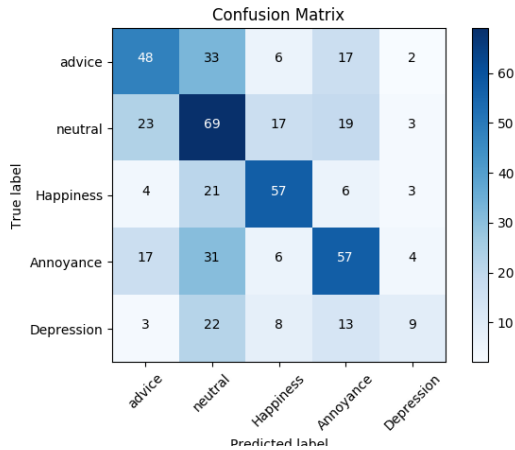


Figure 5. Confusion matrix for CNN

Again we can see that CNN outperforms two other models. The ROC curve generated for the CNN model implemented considering two classes is depicted in Figure 6.

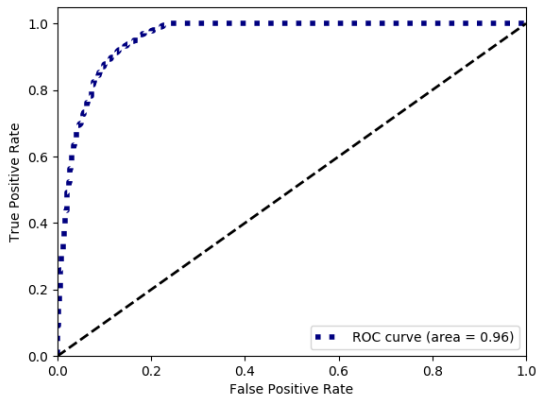


Figure 6. ROC curve for CNN (2 Classes)

But the fact is the overall result is not that satisfactory and not pass the standard and it is obvious because the used dataset is not that large. In a dataset like this classification according to 5 classes is a Panglossian task. So the dataset was decided to be divided into 2 classes; Happiness and Sadness. The same models are implemented in the newly divided dataset. In Table 3 we will find F1 score, precision, and recall of all the three models.

Table 3. Performance measurement of all models (2 classes)

Model	F1 score	Precision	Recall
CNN	0.7557	0.7598	0.7530
LSTM	0.7117	0.7093	0.7168
DNN	0.6867	0.6862	0.7048

The confusion matrix calculated for the CNN model implemented considering two classes is given in figure 7.

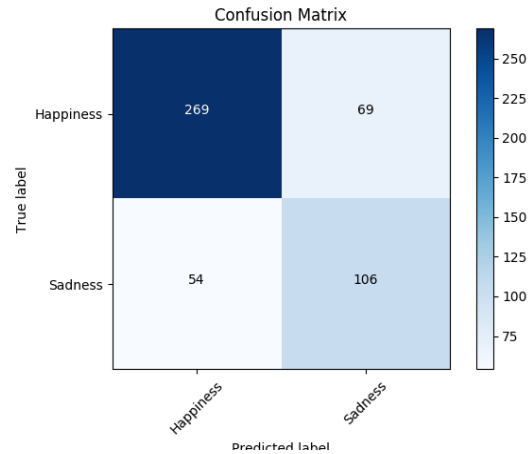


Figure 7. Confusion matrix for CNN (2 classes)

Accuracy of all the models for both 5 and 2 classes experiments are given in figure 8.

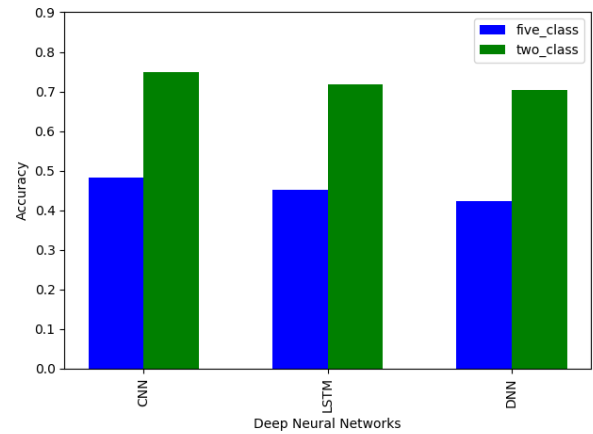


Figure 8. Accuracy comparison of all models

It clearly shows that the performance of all the models make a huge jump when they are implemented on less number of classes. It happens because the number of data per class increases for less number of classes. And this also indicates that our model would work fine also in five classes dataset if the number of data was large enough.

5.1 HINDI DATASETS

Since the dataset was not that large enough to prove the performance of the models, the models were further implemented on two Hindi datasets. That time those models also performed consistently and with better accuracy than the CNN-SVM approach done by this work [14]. The accuracies found by this work on 'HS1' and 'HS2' were **57.34** and **44.88**. We have found CNN outperformed the existing work [14]. Both the F1 scores and accuracies achieved by CNN are given in Table 4.

Table 4. F1 and accuracy measures on Dataset of Hindi language

Model	F1 (HS1)	F1 (HS2)	Accuracy (HS1)	Accuracy (HS2)
CNN	0.65	0.47	67.25	54.10

The ROC curves for the HS1 and HS2 are given in Figure 9 and Figure 10 respectively.

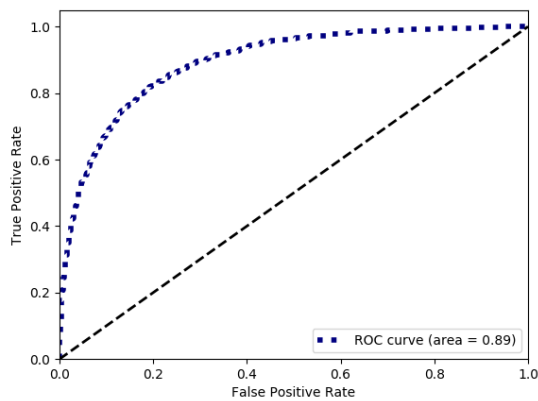


Figure 9. ROC curve for dataset HS1

6. CONCLUSION

Sentiment analysis is a very active research area of NLP. It has a great impact on politics, business, and other social sectors. In this work, different deep neural network architectures are analyzed for sentiment analysis from comments on Bengali Sports News. The type of classification was different here. Authors were more specific about emotions hidden inside a sentence. And a promising model has been established that can classify a sentence according to hidden emotion. For checking the consistency of the model it is also tested using two datasets of Hindi language. It is found that the model worked fine also on those datasets.

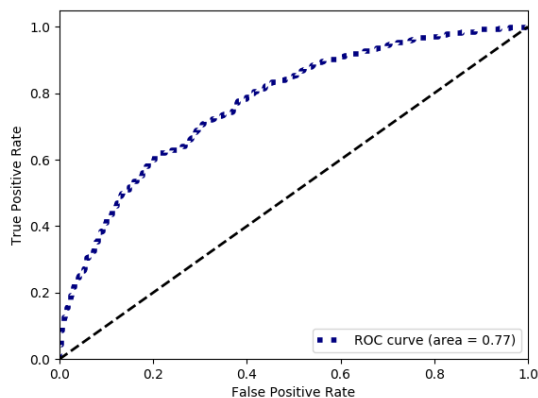


Figure 10. ROC curve for dataset HS2

7. REFERENCES

- [1] Das, D., and Bandyopadhyay, S. 2010. Finding emotion holder from Bengali blog texts -An unsupervised syntactic approach. PACLIC 24 - Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation.
- [2] Sarkar, Iqbal, A., Pavel, D. S. H., and Khan, M. 2007, Automatic Bangla corpus creation. Center for research on Bangla language processing (CRBLP), BRAC University.
- [3] Hasan, Muhammad, F., UzZaman, N., and Khan, M. 2007. Comparison of different POS Tagging Techniques (N-Gram, HMM and Brill's tagger) for Bangla. Advances and Innovations in Systems, Computing Sciences and Software Engineering.
- [4] Majumder, K. M. Y. A., Islam, M. Z., UzZaman, N., and Khan, M. Analysis of and Observations from a Bangla news Corpus. Center for Research on Bangla Language Processing, BRAC University, Dhaka, Bangladesh.
- [5] Nabi, M. M., Altaf, M. T., and Ismail, S. 2016. Detecting Sentiment from Bangla Text using Machine Learning Technique and Feature Analysis International Journal of Computer Applications (0975 – 8887) Volume 153 – No 11.
- [6] Ghosal, T., Das, S. K., and Bhattacharjee, S. Sentiment Analysis on বাংলা রাশিফল (Bengali Horoscope) Corpus.
- [7] Chowdhury, S., and Chowdhury, W. 2014. Performing sentiment analysis in Bangla microblog posts. Int. Conf. Informatics, Electron. Vision, ICIEV
- [8] Ghosal, T., Das, S. K., and Bhattacharjee, S. 2016. Sentiment analysis on (Bengali horoscope) corpus. 12th IEEE Int. Conf. Electron. Energy, Environ. Commun. Comput. Control (E3-C3), INDICON 2015, pp. 1–6.
- [9] Trivedi, M., Soni, N., Sharma, S., and Nair, S. 2015. Comparison of Text Classification Algorithms. International Journal of Engineering Research & Technology (IJERT). vol. 4, no. 2, pp. 334–336.
- [10] Mishu, S. Z., and Rafiuddin, S. M. 2016. Performance Analysis of Supervised Machine Learning Algorithms for Text Classification. 201619th Int. Conf. Comput. Inf. Technol., pp. 409–413.
- [11] Paul, Kumar, A., and Shill, P. C. 2016. Sentiment mining from Bangla data using mutual information. Electrical, Computer & Telecommunication Engineering (ICECTE), International Conference on. IEEE, 2016.
- [12] Islam, M. S., Islam, M. A., Hossain, M. A., and Dey, J. J. 2016. Supervised approach of sentimentality extraction from bengali facebook status. In Computer and Information Technology (ICCIT), 2016 19th International Conference on, pp. 383-387. IEEE.
- [13] Islam, M. S., Al-Amin, M., and Uzzal, S. D. 2016. Word embedding with hellinger PCA to detect the sentiment of bengali text. Computer and Information Technology (ICCIT), 2016 19th International Conference on. IEEE.
- [14] Akhtar, M. S., Kumar, A., Ekbal, A., and Bhattacharyya, P. 2016. A hybrid deep learning architecture for sentiment analysis. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 482-493.