Extracting Protests from News using LSTM Models with different Attention Mechanisms

D. Thenmozhi, Chandrabose Aravindan,
Abishek Shyamsunder, Adithya Vishwanathan and Akash Kumar Pujari

Department of CSE, SSN College of Engineering, Chennai
{theni_d,aravindanc}@ssn.edu.in
abishekshyamsunder@gmail.com
{adithya16002,akash16007}@cse.ssn.edu.in

Abstract. Extracting Protests from news is very useful because it helps in the early identification and subduing of contentious events and controlling violent public outbreaks. This also helps in the study of social sciences on the difference in protest types and their expression across countries. In this paper, a deep learning approach is presented for classifying input documents developed using data from one country and is exercised on data from others. Long Short-Term Memory(LSTM) a variant of Recurrent Neural Network(RNN) is used to implement our approach for the text classification. Models were created for 2 tasks in specific, the first aimed at identifying if a given news document led to a protest or not while the latter predicted if or not an input sentence contained an event trigger. The data given by CodaLab - CLEF 2019 Lab ProtestNews was used to develop and evaluate the performance of the models. The documents and sentences were assigned labels 'protest'/'non-protest' and 'event-trigger'/'not-event-trigger'. A total of 4 LSTM models were implemented with different attention mechanisms, and the ones implementing Scaled-Luong and Bahdanau obtained higher overall F1 macro scores of 0.3290 and 0.3682 on the final evaluation test set respectively.

Keywords: Text Classification · Information Extraction · Long Short Term Model · Recurring Neural Networks · Attention Mechanism

1 Introduction

Text classification refers to the process of assigning tags or categories to textual data according to its content. It is one of the fundamental tasks of Natural Language Processing (NLP) which involves the application of computer techniques to the analysis and synthesis of natural language. Information Extraction, a field strongly associated with the aforementioned topics is the automated retrieval of specific information from a body of text. Using these in unison, the CLEF 2019

Lab ProtestNews TASK 1 aims at classifying news documents as 'protest' or 'non-protest' given raw data and TASK 2 aims at classifying a sentence as one containing an event-trigger or not given the sentence and the news article containing the sentence and Task 3 the extraction of various information from a given event sentence such as location, time and participation of the event [6].

This paper focuses on the use of LSTM, an artificial RNN architecture with different attention mechanisms to perform efficient binary classification of input data and assign class labels for TASK 1 and TASK 2 of CLEF 2019 Lab ProtestNews, functioning along lines similar to that of the novel work published by Peng Zhou et al. [18], which employs Bidirectional LSTM as well as Two-dimensional Max Pooling for the purpose of text classification.

The main challenge is to extract various event information from news articles across multiple countries around the globe. The events considered mainly focus on politics and social movements across the globe collected from Indian and Chinese news articles written in English. Neural Networks can be used for addressing these kinds of problems, but the ability of RNNs to outperform traditional models in applications such as Discriminative Keywords Spotting, evidenced by the works of S Fernandez et al. [4], favours the use of the latter. RNNs use the most recent information to perform the present task. Classifying news articles, however, requires long term dependencies and the vulnerablity of RNNs to hold only the most recent information is overcome by the use of LSTM. As seen by the successful employment of LSTM for Statistical Machine Translation in the paper authored by Kyunghyun Cho et al. [2], they are a special category of RNNs designed specifically to solve the long-term dependency problem.

2 Literature Survey

This section serves to analyse the wide array of published literature and helps in getting an insight into the various models adopted to tackle the problem at hand.

Research works [17, 12, 15, 5] present various machine learning approaches for detecting violence from given textual data. These vary from rule mining [17] to sub-graph pattern mining [12].

Sudha Subramani et al. [15], although focusing on detection of domestic violence, view the problem as one of binary classification much similar to the approach used in this paper. Other studies [7, 10] bring to light different deep learning approaches for the classification of textual data. Madisetty Sreekanth and Desarkar Maunendra Sankar [10] in particular, employ a Convolutional Neural Network-LSTM, which has maximum correlation with the model adopted at present. Mazhar Iqbal Rana et al. [13] provide a comparison of various classifiers in context of short text aiding in sound judgement of baseline establishment.

Big Data Analytics and Probabilistic methods [3, 11] can also be used as alternatives to those mentioned above.

Even though many papers [14, 10, 11] concentrate on data obtained from various social media forums, the concept can be migrated to news articles while taking into consideration the greater length and diversity of news instances. The four models listed in section 4 have established their efficacy through various works published [8, 16, 1, 9], differing from one another along the attention mechanism used. These in their basic form consist of two components, an encoder which computes the representation of each sentence and a decoder which decomposes the conditional probability.

3 Proposed Methodology

This model classifies news documents as protest and non-protest news, event-trigger or not event-trigger and it extracts various information from a given sentence such as location, time and participants of an event. To meet the objectives, the following are done

- Gather news articles
- Preprocessing of news articles
- Grouping and Indexing of news articles
- Feature selection from news articles
- Labelling the news with the help of the model

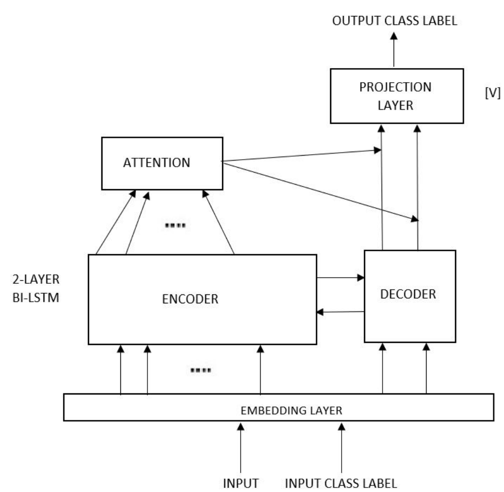

Fig. 1. System Architecture.

Text pre-processing is the initial and important step in news article classification process. pre-processing step makes sure that the learning model receives only the data that are considered useful and the noisy data is removed from the text. The data comes from various sources and ought to be cleaned before it is fed into the model. The noisy data or useless information which includes punctuation marks, non-alpha numeric characters etc. are first removed from the text with the use of regular expressions. A dictionary of unique words is created, and these unique words help in embedding the input sequence to the model to do the classification.

The objects of the training file are separated from one another by using 'newline' as the delimiter. A multi-layer RNN with LSTM as a recurrent unit are used to build a deep neural network model for finding the classified labels. Several layers are used to build the deep neural network namely, embedding layer, encoding-decoding layers, projection layer and loss layer. The weight vectors are learnt from the input sentences and label input based on their vocabulary using the embedding layer. 2-layered bi-directional LSTM is used as hidden layers to perform the encoding and decoding using the embeddings of input sentences or paragraphs. This is shown in Fig 1.

4 Implementation

The current methodology adopted for the TASK1 and TASK2 of CLEF 2019 Lab ProtestNews is implemented using Tensorflow. The data provided, belonging to 4 unique sets namely Development, Training, Testing and Testing_China is presented in Table 1.

Table 1. Data Set for ProtestNews Task1 and Task2

Languages	No. of Instances			
	Development	Training	Test	China_test
Task1*Document*	457	3430	687	1801
Task2*sentence*	663	5885	1107	1235

A total of 4 models [8, 16, 1, 9] have been implemented by using the deep learning approach mentioned in section 3. The 4 models implemented all belong to seq2seq class of algorithms and vary from one another along the attention mechanism used

- Model1- 2 Layered, bi-directional LSTM, with Normed-Bahdanau attention
- Model2- 2 Layered, bi-directional LSTM, with Scaled-Luong attention
- Model3- 2 Layered, bi-directional LSTM, with Bahdanau attention
- Model4- 2 Layered, bi-directional LSTM, with Luong attention

5 Results

A total of 4 models were submitted and were evaluated by the CLEF 2019 Lab ProtestNews Evaluation engine. The test data provided was classified into two sets, the first belonging to news articles from India and the latter containing articles from China, labelled Set 1 for China-Task1, Set 2 for China-Task2, Set 3 for India-Task1 and Set 4 for India-Task2. The Table 2 shows the various F1 macro scores obtained by the models in the final evaluation stage.

Tasks	Model1			Model2			Model3			Model4		
Task1	Set1	Set3	Avg.	Set1	Set3	Avg.	Set1	Set3	Avg.	Set1	Set3	Avg.
	0.027	0.123	0.075	0.134	0.263	0.199	0.158	0.389	0.274	0.233	0.357	0.295
Task2	Set2	Set4	Avg.	Set2	Set4	Avg.	Set2	Set4	Avg.	Set2	Set4	Avg.
	0.015	0.026	0.021	0.370	0.547	0.458	0.357	0.567	0.462	0.196	0.038	0.117
Average	0.048			0.329			0.368			0.206		

Table 2. Final Evaluation Results

Clearly Model 3 for which the 2 Layered, bi-directional LSTM, with Bahdanau attention was adopted produced the best results of the lot with a F1 macro score of 0.3682.

6 Perspectives for future work

From Table 2 given under results, undoubtedly Model3 outperforms its peers for the test data given under Task2. Thus, with the scope of the current study under consideration, adopting the input structure of the training data of Task2 (i.e. sentence wise rather than Document wise) and administering the same in the implementation phase of Task1 may provide better results, thus increasing the overall accuracy. However, there may be two approaches that must be juxtaposed for the post processing of predictions outputted by the model. These are as given below

- At-least-one approach: If any of the sentences belonging to a paragraph/news-article is classified as class 1 (I.e. classified under 'protest') then the whole paragraph/news-article can be classified as class 1.
- Vote-based approach: Each of the sentences can be considered to voting for either class 1 or class 0 when classified. Finally, the whole paragraph/news-article can be classified based on which class obtained the maximum votes.

7 Conclusion

We have developed deep learning models for information extraction and text classification using data from one country and tested them on data from different countries. RNN-LSTM models were used to classify documents and sentences to the labels 'protest'/'non-protest' and 'event-trigger'/'not-event-trigger' respectively. The 4 models implemented differed along the attention mechanism used. The various F1 macro scores obtained were 0.0481, 0.3290, 0.3682 and 0.2064 respectively. The performance may be further improved by the effective splitting of Document data at the pre-processing stage and applying appropriate integration logic of the output at the post-processing stage.

References

1. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
2. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
3. Doyle, A., Katz, G., Summers, K., Ackermann, C., Zavorin, I., Lim, Z., Muthiah, S., Butler, P., Self, N., Zhao, L., et al.: Forecasting significant societal events using the embers streaming predictive analytics system. Big Data **2**(4), 185–195 (2014)
4. Fernández, S., Graves, A., Schmidhuber, J.: An application of recurrent neural networks to discriminative keyword spotting. In: International Conference on Artificial Neural Networks. pp. 220–229. Springer (2007)
5. Hammer, H.L.: Detecting threats of violence in online discussions using bigrams of important words pp. 319–319 (2014)
6. Hürriyetoğlu, A., Yörük, E., Yüret, D., Yoltar, Ç., Gürel, B., Duruşan, F., Mutlu, O.: A task set proposal for automatic protest information collection across multiple countries. In: European Conference on Information Retrieval. pp. 316–323. Springer (2019)
7. Kowsari, K., Brown, D.E., Heidarysafa, M., Meimandi, K.J., Gerber, M.S., Barnes, L.E.: Hdltex: Hierarchical deep learning for text classification. In: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). pp. 364–371. IEEE (2017)
8. Luong, M.T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025 (2015)
9. Luong, T., Brevdo, E., Zhao, R.: Neural machine translation (seq2seq) tutorial. 2017. URL: https://www. tensorflow. org/tutorials/seq2seq (17.02. 2018) (2017)
10. Madisetty, S., Desarkar, M.S.: Aggression detection in social media using deep neural networks. In: Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018). pp. 120–127 (2018)
11. Muthiah, S., Huang, B., Arredondo, J., Mares, D., Getoor, L., Katz, G., Ramakrishnan, N.: Planned protest modeling in news and social media. In: Twenty-Seventh IAAI Conference. pp. 3920–3927 (2015)
12. Qiao, F., Wang, H.: Computational approach to detecting and predicting occupy protest events. In: 2015 International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI). pp. 94–97. IEEE (2015)
13. Rana, M.I., Khalid, S., Akbar, M.U.: News classification based on their headlines: A review. In: 17th IEEE International Multi Topic Conference 2014. pp. 211–216. IEEE (2014)
14. Ranganath, S., Morstatter, F., Hu, X., Tang, J., Wang, S., Liu, H.: Predicting online protest participation of social media users. In: Thirtieth AAAI Conference on Artificial Intelligence. pp. 208–214 (2016)
15. Sudha Subramani, Huy Quan Vu, H.W.: Intent classification using feature sets for domestic violence discourse on social media pp. 129–136 (2017)
16. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. pp. 3104–3112 (2014)
17. Wueest, B., Rothenhäusler, K., Hutter, S.: Using computational linguistics to enhance protest event analysis pp. 1–27 (2013)

18. Zhou, P., Qi, Z., Zheng, S., Xu, J., Bao, H., Xu, B.: Text classification improved by integrating bidirectional lstm with two-dimensional max pooling. arXiv preprint arXiv:1611.06639 (2016)