Political News Bias Detection using Machine Learning

Minh Vu

Earlham College - Department of Computer Science

801 National Road West - Richmond, IN 47374

Introduction

The political climate in the U.S. is sharply divided into two major ideologies: liberal and conservative. In the era of massive social media usage, such ideologies are disseminated at an unprecedented rate. People with certain political opinions get fed information supporting the same political viewpoints, creating an ideological bubble that hinders them from a subjective, neutral information environment[1].

This study aims to use Multilayer Perceptron (MLP), a new, promising neural network used in Natural Language Processing (NLP), to solve the current problem of political bias detection, contributing to the existing body of works that use neural networks to detect political ideologies. It also implements an application (a Google Chrome extension) which can utilize such framework to perform real-time political ideologies classification.

Methodology

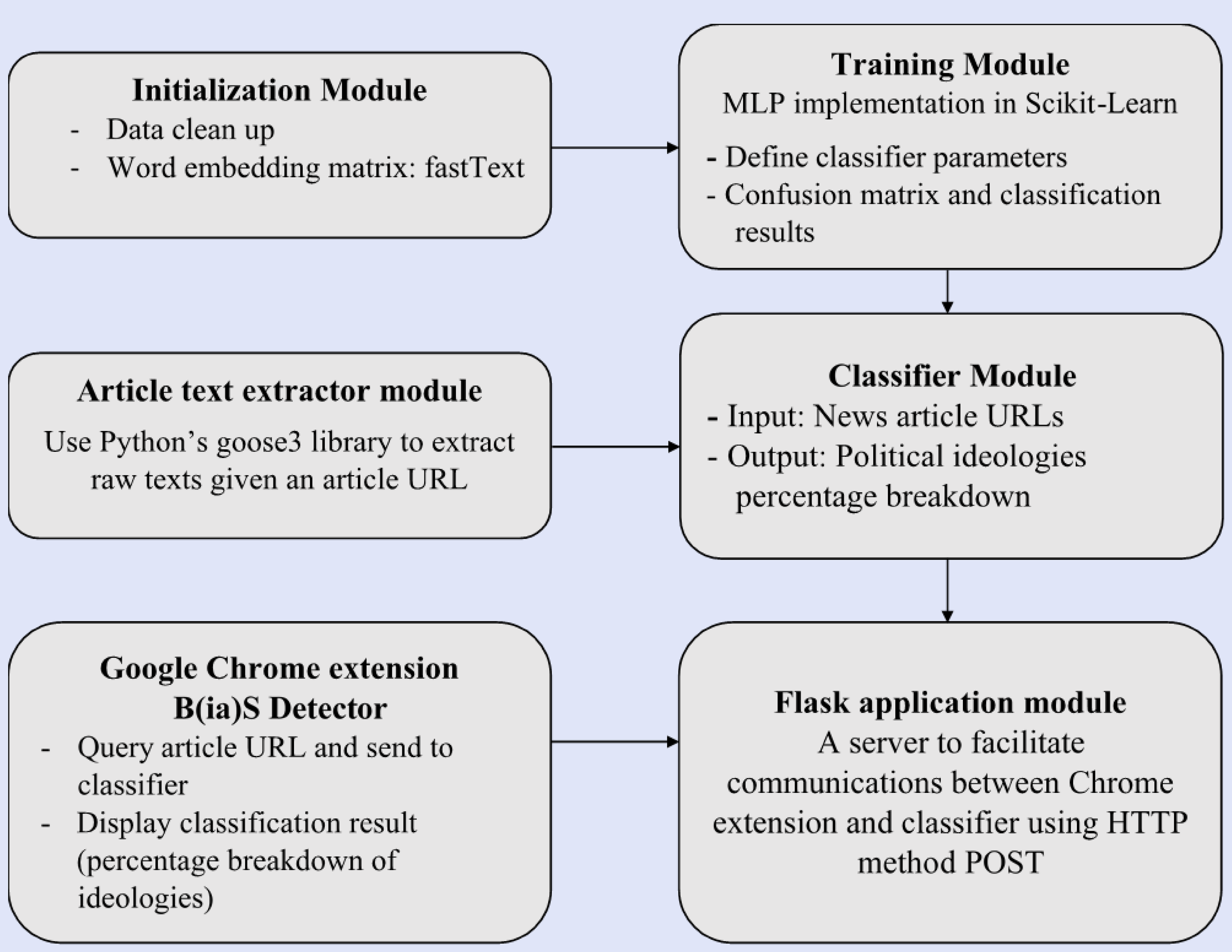

Figure 1. Framework of political news bias detection system

The political news bias detection consists of six main components: Initialization module, training module, classifier module, article text extractor module, Flask application module, and Google Chrome extension (see Fig. 1).

The primary dataset for training the MLP classifier is "The Ideological Books Corpus" (IBC), which consists of 4,062 sentences labeled for political ideology[2]. The IBC is divided into two sets, one used for training and one used for testing, with ratio 3:1.

Results

The first attempt to train the classifier uses the default MLP parameters in Scikit-learn with three hidden layers, each contains ten neurons, the rectifier activation function, Adam optimization algorithm (which works well with large datasets), constant learning rate and 1000 maximum iterations. The F1 score of this training is lower as compared to the model developed by Misra et al.[3] (see Fig. 2).

	Precision	*Recall*	*F1 score*	*Number of sentences*
Conservative	59%	53%	56%	1572
Liberal	64%	66%	65%	1925
Neutral	77%	81%	79%	2159
Average	**68%**	**68%**	**68%**	**5656**

Figure 2. Result of first classifier training

The initial model was refined in four ways:
- Mini-batch gradient descent was employed to take advantage of a matrix-matrix product along with improved efficiency of having all the training data in memory[4].
- Warm start was used, which reuses the aspects of the model learned from the previous parameter value.
- Early stopping was incorporated to prevent overfitting of the model.
- Finally, four hidden layers were used with 500, 20, 20, and 20 neurons, respectively. The new model showed improved performance compared to the previous model (see Fig. 3). It is important to note that the performance was improved with an increase in the size of the first hidden layer but not with subsequent hidden layers.

	Precision	*Recall*	*F1 score*	*Number of sentences*
Conservative	77%	73%	75%	1502
Liberal	81%	79%	80%	1896
Neutral	84%	88%	86%	2258
Average	**81%**	**81%**	**81%**	**5656**

Figure 3. Result of final classifier training

With these results, I also developed a Google Chrome extension, which uses the classifier to provide real-time news bias detection using word-level as well as sentence-level classification (see Fig. 4). The task does not yield promising results, however. Sentence-level and word-level classifications of 20 articles from opinionated news sites both showed "100% Neutral" classification results.

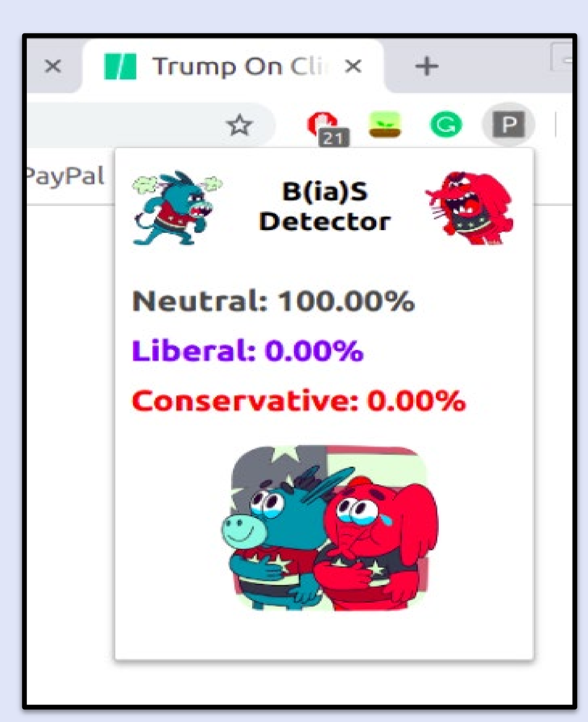

Figure 4. Classification result of real news articles

Discussion

This study has explored a different technical approach to the problem of political bias detection, using a Multilayer Perceptron model, implemented by Python's machine learning library *scikit-learn*. The results of the classification task show that, with a proper set of parameters, the MLP model outperforms the recurrent neural network (RNN) model by Misra et al.[3] on the IBC dataset (F1 scores are 81% and 72%, respectively). However, real-time political news classification (through Chrome extension) did not yield positive results. Both sentence-level and word-level classification of 20 articles from different opinionated news sites, yielded "100% Neutral" results. These outcomes may be attributed to several reasons:
- News sites are reporting with more structurally neutral sentences (while the overall article can still be biased), which neither a sentence-level nor word-level classifier could identify.
- The use of metaphors and negative phrases are not captured by the classifier.
- The MLP classifier does not take into consideration relations between different words and phrases.

Future Works

There are several possible directions which future works on this project can follow:
- Classify sentences based on their semantic structures in order to capture negation and metaphors.
- Utilize larger labeled datasets for better training results.
- Experiment with other word embedding matrices, such as word2vec, GloVe, etc.

References

1.Matthew Carey. 2016. How Donald Trump and Hiillary Clinton are changing the social media game. Los Angeles Daily News (2016).
2.Mohit Iyyer, Peter Enns, Philip Resnik, and Jordan Boyd-Graber. 2014. TheIdeological Books Corpus. https://www.cs.umd.edu/~miyyer/ibc/index.html.(2014)
3.Arkajyoti Misra and Sanjib Basak. 2017. Political Bias Analysis. (2017).
4.Yoshua Bengio. 2012. Practical recommendations for gradient-based training of deep architectures. CoRRabs/1206.5533 (2012)

Acknowledgements

I would like to express my gratitude towards Dr.Ajit Chavan and Dr. Xunfei Jiang for their support, guidance and feedback throughout this project. I would also like to thank Ahsan Khoja and Rei Rembeci for their invaluable help in implementing the Chrome extension and writing up this study.