Assessing the Effectiveness of Affective Lexicons for Depression Classification

Noor Fazilla Abd Yusof, Chenghua Lin and Frank Guerin

Computing Science, University of Aberdeen, Aberdeen UK
{noorfazilla.yusof,chenghua.lin,f.guerin}@abdn.ac.uk

Abstract. Affective lexicons have been commonly used as lexical features for depression classification, but their effectiveness is relatively unexplored in the literature. In this paper, we investigate the effectiveness of three popular affective lexicons in the task of depression classification. We also develop two lexical feature engineering strategies for incorporating those lexicons into a supervised classifier. The effectiveness of different lexicons and feature engineering strategies are evaluated on a depression dataset collected from LiveJournal.

Keywords: Depression analysis, affective lexicon, language analysis

1 Introduction

Depression is one of the most common mental disorders that can affect people of all ages. It is the leading cause of disability and requires significant health care cost to treat effectively [1]. Compared to the traditional clinical consultation, mental health studies based on social media present several advantages [2]. For instance, social media sites provide great venues for people to share their experiences, vent emotion and stress, which are useful for medical practitioners to understand patients' experiences outside the controlled clinical environment. In addition, information captured during clinical consultation generally reflects only the situation of the patient at the time of care. In contrast, data collected from social media is dynamic, thereby providing opportunities for observing and recognising critical changes in patients' behaviour [2].

There is a large body of work using social media for depression related studies, e.g., predicting postpartum depression in new mothers [3], identifying depression-indicative posts [4], and analysing the causes of depression [5]. These works normally leverage various features for the depression classification task such as lexical features (e.g., affective lexicons and linguistic styles), behavioural markers (e.g. time-posting), and social-networking features (e.g. engagement, number of followers). Among these features, lexical features are the most commonly used one. However, their effectiveness in supporting the task of depression classification is relatively unexplored.

In this study, we investigate the effectiveness of three popular affective lexicons in the task of depression classification, namely, ANEW [6], MPQA [7], and SentiWordNet [8]. These lexicons are incorporated into a Multinomial Naive

Bayes (MNB) classifier via two different lexical feature engineering strategies, i.e., (1) include a data instance (i.e., a post) for training only if it contains as least one word from the lexicon, and (2) constrain the training set feature space to the feature space of the lexicon. We test the effectiveness of different lexicons and feature engineering strategies based on a depression dataset collected from LiveJournal. Experimental results show that SentiWordNet outperforms the other lexicons in general, and gives the best result. SentiWordNet achieved 84.4% accuracy using feature engineering Strategy 1, which is significantly better than the baseline (MNB without incorporating any sentiment lexicons).

2   Methodology

In this section, we describe our approach for detecting depressive text by combining a supervised learning model, namely Multinomial Naive Bayes (MNB) [9], with different lexical feature engineering techniques. Our hypothesis is that by leveraging external affective lexicon resources with appropriate feature engineering techniques, the performance of a MNB supervised classifier in predicting depressive text can be significantly enhanced. Specifically, we have explored two lexical feature engineering strategies as detailed below.

– **Strategy 1**: Include a data instance (e.g., a post) for training only if it contains at least one word from the lexicon. This strategy filters out posts that are potentially not relevant to depression.
– **Strategy 2**: Constrain the training set feature space to the features of the affective lexicon. This strategy significantly reduces the feature space of the training set by dropping the words that do not appear in the affective lexicon.

   Please note that we only apply the feature engineering strategies to the training set in order to perform a fair comparison to the baseline, i.e., the original MNB without incorporating any affective lexicons. The test set for the baseline model and our models are identical. Our feature engineering strategies are simple, but are adequate to achieve the main goal of this paper: i.e. to assess the effectiveness of various affective lexicons for depression classification. The feature engineering strategies provide a natural means for incorporating lexicon resources into a MNB supervised classifier.

3   Experimental Setup

Dataset.   We conduct our experiment based on a real-world dataset collected from LiveJournal[1], which consists of 29,686 depressive posts and 87,830 normal (non-depressive) posts. The depressive posts are collected from the depression communities (i.e., identified by searching communities whose name contains the keyword *depress* and its inflections, e.g., *depressed*, *depression*), whereas the normal posts are randomly selected from communities not related to depression. The total number of posts in the dataset is 117,516.

---

[1] http://www.livejournal.com

Table 1: Statistics of Affective Lexicon.

	ANEW	MPQA	SentiWordNet
# of matched affective words	2,430	5,781	2,699
Lexicon coverage	4%	9%	4%

Table 2: Classification Performance of MNB with Different Lexical Feature Engineering Strategies.

Models	Strategy 1				Strategy 2			
	Pre.	Rec.	F1	Acc.	Pre.	Rec.	F1	Acc.
Baseline	63.6	79.8	70.0	81.8	63.6	79.8	70.0	81.8
ANEW	64.1	78.7	69.8	81.9	64.9	57.9	60.3	80.3
MPQA	64.2	78.6	69.9	82.0	60.1	30.5	40.3	77.2
SentiWordNet	68.4	81.3	73.5	84.4	68.8	49.3	57.0	81.0
Combined	69.2	82.1	75.1	**85.3**	70.2	53.2	60.5	**82.8**

Affective lexicons. We have investigated three affective lexicons, i.e., ANEW, MPQA[2] and SentiWordNet[3]. ANEW is a set of 2,476 words that are rated by three emotional dimension (valence, arousal, and dominance). SentiWordNet is a lexical resource that contains 7,247 words which are assigned three sentiment scores (positivity, negativity, and objectivity). MPQA contains 7,319 words for opinions and other private states, e.g., beliefs, emotions, speculations, etc.

Setting. We ran our experiments using the Multinomial Naive Bayes (MNB) algorithm in Weka as it is computationally efficient and often provides competitive performance for text classification problems [9]. Also note that we randomly split the data into 80-20 fractions for training and testing. We repeat the process 5 times and report the averaged results for all models and the baseline.

4 Results

We analyse the statistics of affective lexicons coverage as shown in Table 1. The top row in Table 1 shows the number of words from an affective lexicon that have appeared in the corpus; and bottom row shows the percentage of words in the corpus that are covered by an affective lexicon. We can see that the statistics for both ANEW and SentiWordNet are quite similar, i.e., around 2,500 words of the lexicon have appeared in the dataset, covering 4% of the words in the dataset. The MPQA lexicon, in contrast, gives the highest coverage of about 9% for the entire dataset.

Table 2 shows the classification performance of affective lexicons over two lexical feature engineering strategies. It can be observed that SentiWordNet outperforms ANEW and MPQA under both feature engineering strategies. In particular, SentiWordNet achieves the best overall accuracy of 84.4% with strategy 1, which is about 2.6% higher than the baseline (i.e., 81.8%). This improvement is significant according to a paired T-test with $p < 0.005$. This result is slightly surprising as one would expect the MPQA lexicon to perform best as it has the highest coverage on the dataset. In addition we did a further experiment using

---

[2] http://mpqa.cs.pitt.edu
[3] http://sentiwordnet.isti.cnr.it/

Support Vector Machines (SVM) with RBF kernel; SentiWordNet performed best with 84.7% accuracy with Strategy 1. By combining all the three affective lexicons together (i.e, see 'Combined' in Table 2), we see a further performance boost 85.3%. When comparing the two different feature engineering strategies, we observe that Strategy 2 gives worse performance than Strategy 1. In fact, none of the models incorporating affective lexicons (apart from the combined model) can outperform the baseline. This might be due to the fact that constraining the training set feature space to the features of the affective lexicon has excluded many useful features, leading to a significant drop in model performance. To conclude, SentiWordNet gives the best performance for depression classification based on our feature engineering Strategy 1.

5   Conclusion

In this paper, we investigate the effectiveness of three popular affective lexicons (i.e., ANEW, MPQA, SentiWordNet) in the task of depression classification. We also develop two lexical feature engineering strategies for incorporating those lexicons into a supervised classifier. Our experimental results show that affective lexicons are useful for depression classification, and SentiWordNet is more effective than the other two affective lexicons under both lexical feature engineering strategies. In the future, we would like to conduct experiments on more datasets and test new feature engineering strategies, e.g., to apply topic models.

Acknowledgments.

References

1. Smith, K., Renshaw, P., Bilello, J.: The diagnosis of depression: Current and emerging methods. J. Compr. Psychiatry. 54, 1–6 (2013)
2. Inkster, B., Stillwell, D., Kosinski, M., Jones, P.: A decade into Facebook where is psychiatry in the digital age. J. The Lancet Psychiatry. 01, (2016)
3. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Characterizing and predicting postpartum depression from shared facebook data. In: CSCW. 626-638 (2014)
4. Resnik, P., Armstrong, W., Claudino, L., Nguyen, T., Nguyen, V., Boyd-graber, J.: Beyond LDA: Exploring Supervised Topic Modeling for Depression-Related Language in Twitter. In: CLPsych. 99–107 (2015)
5. Yusof, N F A., Lin, C., Guerin, F.: Analysing the Causes of Depressed Mood from Depression Vulnerable Individuals. In: DDDSM Workshop at IJCNLP. 9–17 (2017)
6. Bradley, M., Lang, P.: Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings. Technical Report C-2. University of Florida. (2010)
7. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: EMNLP-HLT, 347-354 (2005)
8. Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: LREC 10, 2200–2204 (2010)
9. Kibriya, A M., Frank, E., Pfahringer, B., Holmes, G.: Multinomial naive bayes for text categorization revisited. J. AI 2004: Advances in Artificial Intelligence (2005)