USING EYE-TRACKING DATA TO CREATE A WEIGHTED DICTIONARY FOR SENTIMENT ANALYSIS: THE EYE DICTIONARY

Gianpaolo Zammarchi[1], Jaromír Antoch[2]

[1] Department of Economics and Business Sciences, University of Cagliari,
(e-mail: gp.zammarchi@unica.it)

[2] Department of Mathematics and Physics, Charles University,
(e-mail: antoch@karlin.mff.cuni.cz)

ABSTRACT: Extracting information from written texts is of paramount importance to many entities (e.g. businesses, public organizations, individuals), but the exponential growth of available data has made this task beyond any single human being or business. Sentiment analysis is a tool to automatically transform the information extracted into knowledge. One of the main challenges is to assess if a text is positive or negative, which can be tackled using a dictionary where each word has a positive or negative associated value and then combining single-words values to express an overall text sentiment. In order to use such lexicon-based approach, we need an existing dictionary or to build a new one. In this work we present a new dictionary for sentiment analysis developed using eye-tracking data to determine the relevance of words and we assess its performances against other existing dictionaries.

KEYWORDS: eye-tracking, sentiment analysis, lexicon, dictionary.

1 Introduction

Sentiment analysis is aimed at classifying texts into sentiments with a polarity (positive or negative) using different approaches. The lexicon-based approach is based on a dictionary, i.e. a base tool where hundreds or thousands of words are associated with a polarity (negative/positive). In order to classify the polarity of a text, each word is searched in the dictionary. If the word is present, the value assigned to that word will contribute to the overall text sentiment (along with the other words present both in the text and in the dictionary). To obtain a single value representative of the whole text a summarizing function (e.g. average or sum) is applied. An important challenge in sentiment analysis is the definition of weights to attribute to words, i.e. to have instruments to define which words should be assigned greater importance. In this sense, the eye tracking technology, which allows to measure the exact position of the eyes during the visualization of texts, images or other visual stimuli, can be of help to understand which words might be able to gain more attention from a reader and are thus potentially more relevant.

Aim of the present method is to develop a new dictionary for sentiment analysis using eye-tracking data as weights to attribute a different relevance to the words in a text, based on the attention they might receive.

2 Materials and methods

2.1 Development of the Eye-dictionary

To develop a dictionary based on eye tracking data, we focus on two main aspects: weights and polarities. Weights have been computed based on the ProvoCorpus, a large corpus including eye tracking data for 55 paragraphs taken from various sources (e.g. news articles, science magazines and public domain works of fiction). Each paragraph was read by an average of 40 participants. Across all texts, eye tracking data in the form of dwell time for each word (i.e. total reading time calculated as the summation of the duration across all fixations on a given word) are available for a total of 2,689 words (1,191 of which are unique). For each word w included in the corpus of eye tracking data, the average dwell time based on the total number of occurrences of the word in the corpus is calculated as in Eq. (1)

$$\frac{1}{n}\sum_{i=1}^{n} d_i^w \tag{1}$$

where *n* is the number of occurrences of a word *w* in the dataset and d^w is the dwell time for the word *w*. The average global dwell time for any word in the dataset is computed as in Eq. (2)

$$\frac{1}{m}\sum_{i=1}^{m} d_i \tag{2}$$

where *m* is the number of all occurrences of all words observed in the dataset and d_i is the dwell time for the occurrence *i* of a word in the dataset. Each weight v for each word *w* is then calculated as the ratio in Eq. (3)

$$v^w = \frac{\frac{1}{n}\sum_{i=1}^{n} d_i^w}{\frac{1}{m}\sum_{i=1}^{m} d_i} \tag{3}$$

and these values have been normalized using the min-max normalization. Polarities are computed using a large dataset of movie reviews including 50,000 texts, labeled as positive and negative reviews (Maas et al., 2011). To assess if a word has a positive or negative polarity, we compute a probability in the form of Eq. (4):

$$P(w_{pos}) = \frac{N_{w_{pos}}}{N_w} \qquad P(w_{neg}) = \frac{N_{w_{neg}}}{N_w} \tag{4}$$

where $P(w_{pos})$ is the probability that the word *w* is positive, $N_{w_{pos}}$ is the number of occurrences of the word *w* in positive labeled texts and N_w is the number of occurrences of the word w. The same computation is made for negatives. Given the probabilities in Eq. (4) we assign a polarity p to each word w as in Eq. (5)

$$p^w = \begin{cases} 1 & if\ P(w_{pos}) > P(w_{neg}) \\ 0 & if\ P(w_{pos}) = P(w_{neg}) \\ -1 & otherwise \end{cases} \quad (5)$$

Therefore, we assign the word w a positive (+1) or negative value (-1) in case $P(w_{pos})$ is greater or lower than 0.5, respectively. If the probability is exactly 0.5 the word w is assigned 0 (neutral). For each word, a final value s is then computed as the product of weights and polarities as in Eq. (6)

$$s^w = v^w \cdot p^w \quad (6)$$

2.2 Assessment of the performance of the Eye dictionary and comparison with existing dictionaries

The performance of the dictionary based on eye tracking data in the classification of sentiment polarity of texts has been assessed using two independent collections of labeled texts: 1,000 consumer reviews from Amazon (McAuley et al., 2013) and 1,000 consumer reviews from Yelp (Yelp dataset). For these texts, the performance of the Eye dictionary in the classification of sentiment polarity is compared with four existing dictionaries: Loughran-McDonald (2,702 words), SentiWordNet 3.0 (20,093 words), SO-CAL Google (3,290 words) and Hu Liu (6,874 words) extracted from the Lexicon package in R (Rinker, 2018). For each text, a polarity value is calculated as the algebraic sum of signed values assigned to each word by a dictionary. Finally, the number of texts correctly classified using the different dictionaries is compared.

3 Results

A total of 1,185 words for which weights and polarities were computed are included in the Eye dictionary (619 positive, 466 negative and 100 neutral). Table 1 shows the performance of the Eye dictionary and four other dictionaries in terms of precision, recall, F1-score and accuracy for the Yelp dataset (similar results were obtained using the Amazon dataset).

The Eye dictionary showed the best precision for positive texts, best recall for negative texts and the second-best accuracy after the Hu Liu dictionary. The Eye dictionary was able to correctly classify a higher number of texts compared to two of the four dictionaries (Loughran and Socal Google) in the Amazon dataset and three of the four dictionaries (Loughran, Sentiword and Socal Google) in the second dataset. Hu Liu was the only dictionary to show a better performance in both datasets.

Overall, all dictionaries only showed a modest performance in this preliminary analysis, which could be improved with the application of rules for handling cases such as presence of negations, amplifiers and downtoners. Notably, the Eye dictionary was able to achieve a performance similar or better compared to most of the other dictionaries even if it includes a much lower number of words.

Table 1. Comparison between Eye dictionary and four other dictionaries

	Eye dictionary		Loughran-McDonald		SentiWord Net		SO-CAL Google		Hu Liu	
	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Precision	0.60	0.55	0.38	0.30	0.54	0.56	0.48	0.42	0.58	0.68
Recall	0.39	0.74	0.46	0.23	0.63	0.46	0.74	0.19	0.81	0.41
F1-score	0.47	0.63	0.41	0.26	0.58	0.51	0.58	0.27	0.67	0.51
Accuracy	0.56		0.35		0.55		0.46		0.61	

4 Conclusions

In this work we present a new sentiment analysis dictionary built by leveraging eye tracking data to assign weights to words based on their ability to gain attention from a reader. To this aim, dwell time is used as a measure of relevance of a word. Future developments include the expansion of the number of words included in the dictionary as well as evaluation of its performance in the classification of text using rules to handle cases in which classification is particularly challenging, such as sentences including negations, amplifiers and downtoners.

References

KOTZIAS, D., DENIL, M., DE FREITAS, N., & SMYTH, P. 2015. From group to individual labels using deep features. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, USA, 597–606.

MAAS, A., DALY, R.E., PHAM, P.T., HUANG, D., NG, A.Y., & POTTS, C. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1 (HLT '11).* ACL, USA, 142–150.

MCAULEY, J.J., & LESKOVEC, J. 2013. Hidden factors and hidden topics: Understanding rating dimensions with review text. *RecSys '13: Proceedings of the 7th ACM conference on Recommender systems*, 165–172.

RINKER, T.W. 2018. lexicon: Lexicon Data version 1.2.1. http://github.com/trinker/lexicon