Semi-Supervised Affective Meaning Lexicon Expansion Using Semantic and Distributed Word Representations

Areej Alhothali aalhotha@cs.uwaterloo.ca

Jesse Hoey jhoey@cs.uwaterloo.ca

David R. Cheriton School
of Computer Science
University of Waterloo
Waterloo, Ontario, N2L3G1

1 Introduction

Sentiment Analysis is a rapidly growing research area that aims to examine humans' emotions/opinion in subjective documents (e.g., product reviews, news articles, social media posts). One of the most popular methods to perform sentiment analysis tasks is to use sentiment lexicons that associate words with their polarity (e.g., positive or negative). These lexicons can be manually compiled (e.g., Harvard General Inquirer (Stone et al., 1966) and Micro-WNOp (Cerini et al., 2007)) or semi-automatically acquired using some labelled words (e.g., SentiWordNet (Baccianella et al., 2010) and MPQA (Wilson et al., 2005)). Several researchers have generated multi-dimensional (*affective meaning*) lexicons that associate words/ concepts with *affective meaning*, a three-dimensional real-value vectors of evaluation, potency, and activity (i.e., valence, dominance, and arousal) (Osgood, 1957; Bradley and Lang, 1999; Warriner et al., 2013). Other researchers have noted that these multi-dimensional models provide a comprehensive and universal representation of human emotions and that a one- or two- dimensional representation is insufficient to represent humans' emotions (Fontaine et al., 2007). The human-coded semantic lexicons are composed of a relatively small set of words, and incur a high cost for the manual annotation. Further, these lexicons do not cover the wide variety of emerging terms (e.g. on the Internet and social media such as *selfie*, *sexting*, or *photobomb*). To overcome this limitation, we propose an approach to extend an affective meaning lexicon using semantic and distributed word representations. Our approach propagates existing sentiment annotations to new words according to word similarity, either according to word semantics or word distributions in large-scale corpora. The results show that using word semantic neural word embedding generates the highest correlations ($\tau = 0.51$) and error rates less than 1.1 with existing labels.

2 Methodology

Label propagation algorithms (Zhu and Ghahramani, 2002; Zhou et al., 2004) rely on the idea of building a similarity graph with labelled (seed words/paradigm words) and unlabelled nodes (words). The labels or scores of the known nodes (words) are then propagated through the graph to the unlabelled nodes by repeatedly multiplying the weight matrix (affinity matrix) against the labels or scores vector. Following the same principle, the graph label propagation algorithm in this paper: 1) creates a set of labeled $L = (X_l, Y_l)$ and unlabelled data points or words $U = (X_u, Y_u)$ where $|U| + |L| = |V|$, V is all the words in the vocabulary set, X is the word, and Y is the sentiment (E, P, A scores) attached to that word; 2) constructs an undirected weighted graph $G = \{E, V, W\}$ where V is a set of vertices (words), E edges, W is an $|V| \times |V|$ weight matrix (where $w_{ij} \geq 0$); 3) Computes the random walk normalized Laplacian matrix $\Delta = D^{-1}W$ (where D is the degree matrix); 4) initializes the labeled nodes/words Y_l with their EPA values, and the unlabeled nodes/words Y_u with zeroes; 4) propagates the sentiment scores to adjacent nodes by computing $Y \leftarrow \Delta Y$ (weighted by a factor α) and clamps the labeled nodes Y_l to their initial values L after each iteration.

We implemented the label propagation algorithm using four different methods of computing affinity matrix and word representations. First, a semantic lexicon-based label propagation (SLLP) in which the graph is built based upon the semantic relationship between words. These semantic features were obtained from WordNet dictionary (WN) (Miller, 1995) and the paraphrase database (PPDB) (Ganitkevitch et al., 2013).

Second, a corpus-based label propagation (CLP) in which vocabulary and weights come from co-occurrence statistics in corpora. The co-occurrence statistics were gathered from the signal media (SM) (Corney et al., 2016) and the North American News (NAN) (Graff, 1995) news corpora. Third, a neural word embedding label propagation (NWELP) that uses two pre-trained word embedding models: skip-gram model (SG) (Mikolov et al., 2013) and Global vector for word representation (GloVe) (Pennington et al., 2014), and fourth, a combination of semantic and distributional methods (semantic neural word embedding label propagation - SNWELP). We use the algorithms to augment a manually-annotated affective dictionary (Warriner et al., 2013). We randomly divided the dictionary (rescaled into $\in [-4.3, +4.3]$) into training-set (5566 words) and testing-set (8349 words). The seed words, which contribute to no more than 1% of all words in each algorithm, are sampled from the training-set, and all results are presented on the testing-set. Words are sampled non-randomly such that they maximally span the EPA space. To compare the induced EPA against the manually annotated EPA, we used Kendall τ rank correlation and mean absolute error (MAE).

The four algorithms just described are all *semi-supervised* in that they only require labels on a small fraction of the training set. In order to ground the results, we used a *supervised* learning algorithm to train a support vector regression (SVR) model on co-occurrence statistics derived from the skip-gram word embedding model (SG) (Mikolov et al., 2013). The supervised method uses *all* the labels in the training set (over 5000 labels).

3 Results

Table 1 shows the results of comparing the induced EPA scores using the label propagation algorithms against their corresponding values in the testing-set. Table 2 shows some of the induced EPA scores and their corresponding values in (Warriner et al., 2013) dataset.

Method	**Corpus**	τ			**MAE**		
		E	P	A	E	P	A
CLP	SM	0.219	0.0263	0.162	1.10	1.09	0.85
	NAN	0.122	0.060	0.084	1.30	1.0	0.99
SLLP	WN	0.388	0.244	0.329	0.91	**0.79**	**0.71**
	PPDB	0.391	0.181	0.309	0.92	0.89	0.79
NWELP	SG	0.437	0.283	0.350	**0.84**	1.08	0.88
	GloVe	0.430	0.113	0.357	1.09	1.07	0.84
SNWELP	PPDB+GloVe	0.434	0.209	0.360	1.09	1.07	0.84
	WN+GloVe	0.445	0.220	0.366	1.07	1.05	0.84
	PPDB+SG	**0.510**	0.284	0.459	1.10	0.97	0.84
	WN+SG	**0.510**	**0.291**	**0.461**	1.10	0.95	0.83
Supervised	SVR	**0.628***	**0.422***	**0.500***	**0.60***	**0.60***	**0.56***

Table 1: The results of the label propagation algorithms in comparison with the ground truth EPA values. Method= the algorithm used for lexicon induction, τ = Kendall's τ correlation and MAE=Mean Absolute Error. CLP, SLLP, NWELP and SNWELP are all *semi-supervised* and use only a small set of labels for **sampled seed words**, where as the Supervised algorithm (bottom row) uses all labels in the training set. The highest scores of the label propagation algorithms are in a **boldface**. The highest scores of all the algorithms are in **boldface***.

Word	**Method**	**Induced EPA**	**True EPA**
injustice	SG+WN	[-1.9, 0.3, -1.7]	[-2.7, 1.6, -1.86]
fearful	SG+WN	[-2.8, -0.1 , -2.0]	[-2.5, 0.5, -2.0]
evil	SG+PPDB	[-2.1, 0.1, -1.5]	[-2.9, 0.7, -1.5]
successful	SG+PPDB	[2.5, -0.6, 2.0]	[2.97, 0.09, 2.9]

Table 2: Some example of the induced EPA and their EPA ratings from Original-EPA-lexicon and the induced EPA values using semantic and neural word embeddings label propagation (SNWELP).

4 Discussion and Conclusion

In this study, we propose a set of semi-supervised graph-based lexicon induction algorithms to expand sentiment lexicons. We found that, with only as few as 50 labeled words, error rates as low as 0.83 and τ correlation as high as 0.51 are possible in some dimensions. Comparing the results across the different affective dimensions (E,P, and A) shows that the rank correlation τ for potency (P) was low in comparison with the scores for evaluation (E) and activity (A) in both the semi-supervised algorithms and the supervised algorithm. While the rank correlation τ of the evaluation (E) scores were the highest in all the algorithms. This would indicate that words with similar word embeddings have a similar evaluation scores.

References

[1] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining." In: *LREC*. Vol. 10. 2010, pp. 2200–2204.

[2] Margaret M Bradley and Peter J Lang. *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Tech. rep. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999.

[3] Sabrina Cerini, Valentina Compagnoni, Alice Demontis, Maicol Formentelli, and G Gandini. "Language resources and linguistic theory: Typology, second language acquisition, English linguistics, chapter Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining". In: *Franco Angeli Editore, Milano, IT* (2007).

[4] David Corney, Dyaa Albakour, Miguel Martinez, and Samir Moussa. "What do a Million News Articles Look like?" In: *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*. 2016, pp. 42–47. URL: `http://ceur-ws.org/Vol-1568/paper8.pdf`.

[5] Johnny RJ Fontaine, Klaus R Scherer, Etienne B Roesch, and Phoebe C Ellsworth. "The world of emotions is not two-dimensional". In: *Psychological science* 18.12 (2007), pp. 1050–1057.

[6] Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. "PPDB: The Paraphrase Database." In: *HLT-NAACL*. 2013, pp. 758–764.

[7] David Graff. *North american news text corpus*. 1995.

[8] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality". In: *Advances in neural information processing systems*. 2013, pp. 3111–3119.

[9] George A Miller. "WordNet: a lexical database for English". In: *Communications of the ACM* 38.11 (1995), pp. 39–41.

[10] Charles Egerton Osgood. *The measurement of meaning*. Vol. 47. University of Illinois Press, 1957.

[11] Jeffrey Pennington, Richard Socher, and Christopher D Manning. "Glove: Global Vectors for Word Representation." In: *EMNLP*. Vol. 14. 2014, pp. 1532–43.

[12] Philip J Stone, Dexter C Dunphy, and Marshall S Smith. "The general inquirer: A computer approach to content analysis." In: (1966).

[13] Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. "Norms of valence, arousal, and dominance for 13,915 English lemmas". In: *Behavior research methods* 45.4 (2013), pp. 1191–1207.

[14] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis". In: *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics. 2005, pp. 347–354.

[15] Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schölkopf. "Learning with local and global consistency". In: *Advances in neural information processing systems* 16.16 (2004), pp. 321–328.

[16] Xiaojin Zhu and Zoubin Ghahramani. *Learning from labeled and unlabeled data with label propagation*. Tech. rep. Technical Report CMU-CALD-02-107, Carnegie Mellon University, 2002.