Predicting Valence-Arousal Ratings of Words Using a Weighted Graph Method

Liang-Chih Yu[1,3], Jin Wang[2,3,4], K. Robert Lai[2,3] and **Xue-jie Zhang[4]**

[1]Department of Information Management, Yuan Ze University, Taiwan
[2]Department of Computer Science & Engineering, Yuan Ze University, Taiwan
[3]Innovation Center for Big Data and Digital Convergence Yuan Ze University, Taiwan
[4]School of Information Science and Engineering, Yunnan University, Yunnan, P.R. China
Contact: lcyu@saturn.yzu.edu.tw

Abstract

Compared to the categorical approach that represents affective states as several discrete classes (e.g., positive and negative), the dimensional approach represents affective states as continuous numerical values on multiple dimensions, such as the valence-arousal (VA) space, thus allowing for more fine-grained sentiment analysis. In building dimensional sentiment applications, affective lexicons with valence-arousal ratings are useful resources but are still very rare. Therefore, this study proposes a weighted graph model that considers both the relations of multiple nodes and their similarities as weights to automatically determine the VA ratings of affective words. Experiments on both English and Chinese affective lexicons show that the proposed method yielded a smaller error rate on VA prediction than the linear regression, kernel method, and pagerank algorithm used in previous studies.

1 Introduction

Thanks to the vigorous development of online social network services, anyone can now easily publish and disseminate articles expressing their thoughts and opinions. Sentiment analysis thus has become a useful technique to automatically identify affective information from texts (Pang and Lee, 2008; Calvo and D'Mello, 2010; Liu, 2012; Feldman, 2013). In sentiment analysis, representation of affective states is an essential issue and can be generally divided into categorical and dimensional approaches.

The categorical approach represents affective states as several discrete classes such as binary (positive and negative) and Ekman's six basic emotions (e.g., anger, happiness, fear, sadness, disgust and surprise) (Ekman, 1992). Based on this representation, various techniques have been investigated to develop useful applications such as deceptive opinion spam detection (Li et al., 2014), aspect extraction (Mukherjee and Liu, 2012), cross-lingual portability (Banea et al., 2013; Xu et al., 2015), personalized sentiment analysis (Ren and Wu, 2013; Yu et al., 2009) and viewpoint identification (Qiu and Jiang, 2013). In addition to identifying sentiment classes, an extension has been made to further determine their sentiment strength in terms of a multi-point scale (Taboada et al., 2011; Li et al., 2011; Yu et al., 2013; Wang and Ester, 2014).

The dimensional approach has drawn considerable attention in recent years as it can provide a more fine-grained sentiment analysis. It represents affective states as continuous numerical values on multiple dimensions, such as valence-arousal (VA) space (Russell, 1980), as shown in Figure 1. The valence represents the degree of pleasant and unpleasant (or positive and negative) feelings, and the arousal represents the degree of excitement and calm. Based on such a two-dimensional representation, a common research goal is to determine the degrees of valence and arousal of given texts such that any affective state can be represented as a point in the VA coordinate plane. To accomplish this goal, affective lexicons with valence-arousal ratings are useful resources but few exist. Most existing applications rely on a handcrafted lexicon ANEW (Af-
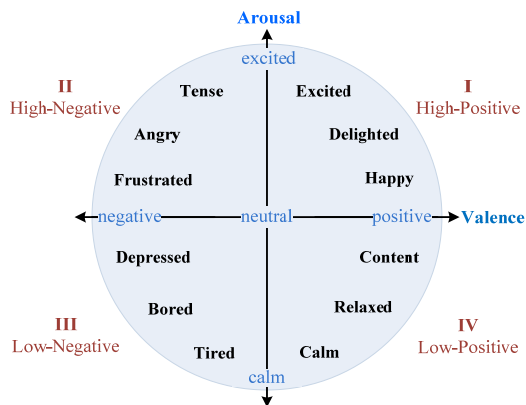
Figure 1. Two-dimensional valence-arousal space.

fective Norms for English Words) (Bradley, 1999) which provides 1,034 English words with ratings in the dimensions of pleasure, arousal and dominance to predict the VA ratings of short and long texts (Paltoglou et al, 2013; Kim et al., 2010). Accordingly, the automatic prediction of VA ratings of affective words is a critical task in building a VA lexicon.

Few studies have sought to predict the VA rating of words using regression-based methods (Wei et al., 2011; Malandrakis et al., 2011). This kind of method usually starts from a set of words with labeled VA ratings (called seeds). The VA rating of an unseen word is then estimated from semantically similar seeds. For instance, Wei et al. (2011) trained a linear regression model for each seed cluster, and then predicted the VA rating of an unseen word using the model of the cluster to which the unseen word belongs. Malandrakis et al. (2011) used a kernel function to combine the similarity between seeds and unseen words into a linear regression model. Instead of estimating VA ratings of words, another direction is to determine the polarity (i.e., positive and negative) of words by applying the label propagation (Rao and Ravichandran, 2009; Hassan et al., 2011) and pagerank (Esuli et al., 2007) on a graph. Based on these methods, the polarity of an unseen word can be determined/ranked through its neighbor nodes (seeds).

Although the pagerank algorithm has been used for polarity ranking, it can still be extended for VA prediction. Therefore, this study extends the idea of pagerank in two aspects. First, we implement pagerank for VA prediction by transforming ranking scores into VA ratings. Second, whereas pagerank assigns an equal weight to the edges connected between an unseen word and its neighbor nodes, we consider their similarities as weights to construct a weighted graph such that neighbor nodes more similar to the unseen word may contribute more to estimate its VA ratings. That is, the proposed weighted graph model considers both the relations of multiple nodes and the similarity weights among them. In experiments, we evaluate the performance of the proposed method against the linear regression, kernel method, and pagerank algorithm on both English and Chinese affective lexicons for VA prediction.

The rest of this paper is organized as follows. Section 2 describes the proposed weighted graph model. Section 3 summarizes the comparative results of different methods for VA prediction. Conclusions are finally drawn in Section 4.

2 Graph Model for VA Prediction

Based on the theory of link analysis, the relations between unseen words and seed words can be considered as a graph, as shown in Figure 2. The valence-arousal ratings of each unseen word can then be predicted through the links connected to the seed words to which it is similar using their similarities as weights. To measure the similarity between words (nodes), we use the word2vec toolkit (Mikolov et al., 2013) provided by Google (http://code.google.com/p/word2vec/).

The formal definition of a graph model is described as follows. Let $G=(V, E)$ be an undirected graph, where V denotes a set of words and E denotes a set of undirected edges. Each edge e in E denotes a relation between word v_i and word v_j in V ($1 \leq i, j \leq n, i \neq j$), representing the similarity between them. For each node v_i, $N(v_i) = \{v_j \mid (v_j, v_i) \in E\}$ denotes the set of its neighbor nodes, representing a set of words to which it is similar. The valence or arousal of v_i, denoted as val_{v_i} or aro_{v_i}, can then be determined by its neighbors, defined as

$$val_{v_i} = (1-\alpha) \cdot val_{v_i} + \alpha \frac{\sum_{v_j \in N(v_i)} Sim(v_i, v_j) \cdot val_{v_j}}{\sum_{v_j \in N(v_i)} Sim(v_i, v_j)},$$

(1)

where α is a decay factor or a confidence level for computation (a constant between 0 and 1), which limits the effect of rank sinks to guarantee convergence to a unique vector. Initially, the valence (or arousal) of each unseen word is assigned to a random value that between 0 and 10. Later, it is iteratively updated using the following formula,
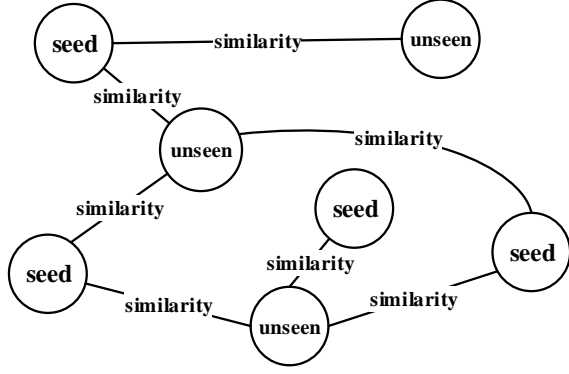
Figure 2. Conceptual diagram of a weighted graph model for VA prediction.

$$val_{v_i}^t = \begin{cases} RandomValue & (t=0) \\ (1-\alpha)\cdot val_{v_i}^{t-1} + \alpha \dfrac{\sum_{v_j \in N(v_i)} Sim(v_i,v_j)\cdot val_{v_j}^{t-1}}{\sum_{v_j \in N(v_i)} Sim(v_i,v_j)} & (t>0) \end{cases} \tag{2}$$

where t denotes the t-th iteration. It is worth noting that the valence (or arousal) of the seed words is a constant in each iterative step. Based on this, the valence (or arousal) of each unseen word is propagated through the graph in multiple iterations until convergence.

To improve the efficiency of the iterative computation, Eq. (2) can be transformed into a matrix notation. Suppose that the vectors,

$$\mathbf{V} = (val_{v_1}, val_{v_1}, \cdots, val_{v_N})^T,$$
$$\mathbf{A} = (aro_{v_1}, aro_{v_1}, \cdots, aro_{v_N})^T$$

are the vectors of the valence-arousal rating of all words (including seed words and unseen words). Matrix

$$\mathbf{S} = \begin{bmatrix} Sim(v_1,v_1) & \cdots & Sim(v_1,v_j) & \cdots & Sim(v_1,v_N) \\ \vdots & & \vdots & & \vdots \\ Sim(v_i,v_1) & \cdots & Sim(v_i,v_j) & \cdots & Sim(v_i,v_N) \\ \vdots & & \vdots & & \vdots \\ Sim(v_N,v_1) & \cdots & Sim(v_N,v_j) & \cdots & Sim(v_N,v_N) \end{bmatrix}$$

is the adjacency matrix of each words, where $Sim(v_i, v_j)$ represents the similarity between words i and j, where $i,j=1, 2, \cdots, N, i \neq j$.

Given two other vectors $\mathbf{I} = (1,1,\cdots,1)^T$ and $\mathbf{D} = (d_1, d_2, \cdots, d_N)^T$, where

$$d_i = \begin{cases} \alpha & if \quad node_i \in \textbf{cand} \\ 0 & if \quad node_i \in \textbf{seed} \end{cases},$$

α is the previously mentioned decay factor. For vectors $\mathbf{A}=(a_1,a_2,\cdots,a_N)^T$ and $\mathbf{B}=(b_1,b_2,\cdots,b_N)^T$, function $\mathcal{M}(\mathbf{A},\mathbf{B})$ and $\mathcal{D}(\mathbf{A},\mathbf{B})$ can be defined as

$$\mathcal{M}(\mathbf{A},\mathbf{B}) = (a_1 \times b_1, a_2 \times b_2, \cdots, a_N \times b_N)^T,$$
$$\mathcal{D}(\mathbf{A},\mathbf{B}) = (a_1 / b_1, a_2 / b_2, \cdots, a_N / b_N)^T.$$

Then, Eq. (2) can be turned into the following matrix format.

$$\mathbf{V}_t = \mathcal{M}[(\mathbf{I}-\mathbf{D})^T, \mathbf{V}_{t-1}] + \mathcal{M}[\mathbf{D}^T, \mathcal{D}(\mathbf{SV}_{t-1}, \mathbf{S}\cdot\mathbf{I})],$$
$$\mathbf{A}_t = \mathcal{M}[(\mathbf{I}-\mathbf{D})^T, \mathbf{A}_{t-1}] + \mathcal{M}[\mathbf{D}^T, \mathcal{D}(\mathbf{SA}_{t-1}, \mathbf{S}\cdot\mathbf{I})] \tag{3}$$

Through the transformation of matrix multiplication, the computation of VA prediction can converge within only a few iterations.

3 Experimental Results

Data. This experiment used two affective lexicons with VA ratings: 1) ANEW which contains 1,034 English affective words (Bradley, 1999) and 2) 162 Chinese affective words (CAW) taken from (Wei et al., 2011). Both lexicons were used for 5-fold cross-validation. That is, for each run, 80% of the words in the lexicons were considered as seeds and the remaining 20% were used as unseen words. The similarities between English words and between Chinese words were calculated using the word2vec toolkit trained with the respective English and Chinese wiki corpora (https://dumps.wikimedia.org/).

Implementation Details. Two regression-based methods were used for comparison: linear regression (Wei et al., 2011) and the kernel method (Malandrakis et al., 2011), along with two graph-based methods: pagerank (Esuli et al., 2007) and the proposed weighted graph model. For both regression-based methods, the similarities and VA ratings of the seed words were used for training, and the VA ratings of an unseen word were predicted by taking as input its similarity to the seeds. In addition, for the kernel method, the linear similarity function was chosen because it yielded top performance. Both graph-based methods used an iterative procedure for VA prediction and required no training. For pagerank, the iterative procedure was implemented using the algorithm presented in (Esuli et al., 2007), which estimates the VA ratings of an unseen word by assigning an equal weight to the edges connected to its neighbor seeds. For the proposed method, the iterative procedure was implemented by considering the word similarity as weights.

Valence	ANEW (English)			CAW (Chinese)		
	RMSE	MAE	MAPE(%)	RMSE	MAE	MAPE(%)
Weighted Graph	1.122	0.812	11.51	1.224	0.904	13.03
PageRank	1.540	1.085	15.69	1.642	1.187	16.84
Kernel	1.926	1.385	19.55	2.028	1.426	20.57
Linear Regression	1.832	1.301	18.61	1.935	1.393	19.66
Arousal	ANEW (English)			CAW (Chinese)		
	RMSE	MAE	MAPE(%)	RMSE	MAE	MAPE(%)
Weighted Graph	1.203	0.894	12.24	1.311	0.966	13.37
PageRank	1.627	1.149	16.48	1.735	1.238	17.51
Kernel	2.007	1.419	20.27	2.118	1.434	21.44
Linear Regression	1.912	1.382	19.33	2.020	1.421	20.46

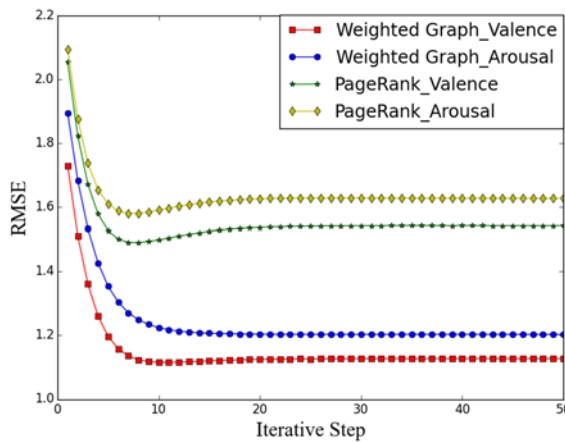Table 1. Comparative results of different methods in VA prediction.

Figure 3. Iterative results of the pagerank algorithm and weighted graph model.

Evaluation Metrics. The prediction performance was evaluated by examining the difference between the predicted values of VA ratings and the corresponding actual values in the ANEW and CAW lexicons. The evaluation metrics included:

- *Root mean square error* (RMSE)

$$RMSE = \sqrt{\sum_{i=1}^{n}(A_i - P_i)^2 \Big/ n}$$

- *Mean absolute error* (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|A_i - P_i|,$$

- *Mean absolute percentage error* (MAPE)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|A_i - P_i|}{A_i} \times 100\%$$

where A_i is the actual value, P_i is the predicted value, and n is the number of test samples. A lower MAPE, MAE or RMSE value indicates more accurate forecasting performance.

Iterative Results of the Graph-based Methods. Figure 3 uses RMSE as an example to show the iterative results of the pagerank and proposed methods. The results show that the performance of both methods stabilized after around 10 iterations, indicating its efficiency for VA prediction. Another observation is that the ultimate converging result of each word is unrelated to the decay factor and the initial random assignment.

Comparative Results. Table 1 compares the results of the regression-based methods (Linear Regression and Kernel) and graph-based methods (PageRank and Weighted Graph). The performance of PageRank and Weighted Graph was taken from results of the 50th iteration. The results show that both graph-based methods outperformed the regression-based methods for all metrics. For the graph-based methods, the proposed Weighted Graph yielded better MAPE performance than PageRank (around 4%), Kernel (around 8%) and Linear Regression (around 7%) on both the ANEW and CAW corpora. The weighted graph model achieved better performance because it predicted VA ratings by considering both the relations of multiple nodes and the weights between them. For the regression-based methods, both Linear Regression and Kernel achieved similar results. Another observation is that the arousal prediction error is greater than that for the valence prediction, indicating that the arousal dimension is more difficult to predict.

791

4   Conclusion

This study presents a weighted graph model to predict valence-arousal ratings of words which can be used for lexicon augmentation in the valence and arousal dimensions. Unlike the equal weight used in the traditional pagerank algorithm, the proposed method considers the similarities between words as weights such that the neighbor nodes more similar to the unseen word may contribute more to VA prediction. Experiments on both English and Chinese affective lexicons show that the proposed method yielded a smaller error rate than the pagerank, kernel and linear regression methods. Future work will focus on extending the VA prediction from the word-level to the sentence- and document-levels.

Acknowledgments

Reference

Carmen Banea, Rada Mihalcea, and Janyce Wiebe. 2013. Porting multilingual subjectivity resources across languages. *IEEE Trans. Affective Computing*, 4(2):211-225.

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

Rafael A. Calvo and Sidney D'Mello. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Trans. Affective Computing*, 1(1): 18-37.

Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion*, 6:169-200.

Andrea Esuli and Fabrizio Sebastiani. 2007. Pageranking wordnet synsets: An application to opinion mining. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL-07)*, pages 424-431.

Ronen Feldman. 2013. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82-89.

Ahmed Hassan, Amjad Abu-Jbara, Rahul Jha, Dragomir Radev. 2011. Identifying the semantic orientation of foreign words. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-11)*, pages 592-597.

Sunghwan Mac Kim, Alessandro Valitutti, and Rafael A. Calvo. 2010. Evaluation of unsupervised emotion models to textual affect recognition. In *Proc. of Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 62-70.

Fangtao Li, Nathan Liu, Hongwei Jin, Kai Zhao, Qiang Yang, and Xiaoyan Zhu. 2011. Incorporating reviewer and product information for review rating prediction. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pages 1820-1825.

Jiwei Li, Myle Ott, Claire Cardie, and Eduard Hovy. 2014. Towards a general rule for identifying deceptive opinion spam. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL-14)*, pages 1566-1576.

Bing Liu. 2012. Sentiment Analysis and Opinion Mining. Morgan & Claypool, Chicago, IL.

Nikos Malandrakis, Alexandros Potamianos, Iosif Elias, and Shrikanth Narayanan. 2011. Kernel models for affective lexicon creation. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech-11)*, pages 2977-2980.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26 (NIPS-13)*, pages 3111-3119.

Arjun Mukherjee and Bing Liu. 2012. Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL-12)*, pages 339-348.

Georgios Paltoglou, Mathias Theunis, Arvid Kappas, and Mike Thelwall. 2013. Predicting emotional responses to long informal text. *IEEE Trans. Affective Computing*, 4(1):106-115.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1-135.

Minghui Qiu and Jing Jiang. 2013. A latent variable model for viewpoint discovery from threaded forum posts. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL/HLT-13)*, pages 1031-1040.

Delip Rao and Deepak Ravichandran. 2009. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-09)*, pages 675–682.

Fuji Ren and Ye Wu. 2013. Predicting user-topic opinions in Twitter with social and topical context. *IEEE Trans. Affective Computing*, 4(4):412-424.

James A. Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267-307.

Hao Wang and Martin Ester. 2014. A sentiment-aligned topic model for product aspect rating prediction. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP-14)*, pages 1192-1202.

Wen-Li Wei, Chung-Hsien Wu, and Jen-Chun Lin. 2011. A regression approach to affective rating of Chinese words from ANEW. In *Proc. of Affective Computing and Intelligent Interaction (ACII-11)*, pages 121-131.

Ruifeng Xu, Lin Gui, Jun Xu, Qin Lu, and Kam-Fai Wong. 2015. Cross lingual opinion holder extraction based on multi-kernel SVMs and transfer learning. *World Wide Web*, 18:299-316.

Liang-Chih Yu, Chung-Hsien Wu, and Fong-Lin Jang. 2009. Psychiatric document retrieval using a discourse-aware model. *Artificial Intelligence*, 173(7-8): 817-829.

Liang-Chih Yu, Jheng-Long Wu, Pei-Chann Chang, and Hsuan-Shou Chu. 2013. Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. *Knowledge-based Systems*. 41:89-97.