Dimensional Sentiment Analysis Using a Regional CNN-LSTM Model

Jin Wang[1,3,4], Liang-Chih Yu[2,4], K. Robert Lai[3,4] and Xuejie Zhang[1]
[1]School of Information Science and Engineering, Yunnan University, Yunnan, P.R. China
[2]Department of Information Management, Yuan Ze University, Taiwan
[3]Department of Computer Science & Engineering, Yuan Ze University, Taiwan
[4]Innovation Center for Big Data and Digital Convergence Yuan Ze University, Taiwan
Contact: lcyu@saturn.yzu.edu.tw

Abstract

Dimensional sentiment analysis aims to recognize continuous numerical values in multiple dimensions such as the valence-arousal (VA) space. Compared to the categorical approach that focuses on sentiment classification such as binary classification (i.e., positive and negative), the dimensional approach can provide more fine-grained sentiment analysis. This study proposes a *regional* CNN-LSTM model consisting of two parts: regional CNN and LSTM to predict the VA ratings of texts. Unlike a conventional CNN which considers a whole text as input, the proposed regional CNN uses an individual sentence as a region, dividing an input text into several regions such that the useful affective information in each region can be extracted and weighted according to their contribution to the VA prediction. Such regional information is sequentially integrated across regions using LSTM for VA prediction. By combining the regional CNN and LSTM, both local (regional) information within sentences and long-distance dependency across sentences can be considered in the prediction process. Experimental results show that the proposed method outperforms lexicon-based, regression-based, and NN-based methods proposed in previous studies.

1 Introduction

Sentiment analysis has been useful in the development of online applications for customer reviews and public opinion analysis (Pang and Lee 2008; Calvo and D'Mello 2010; Liu 2012; Feldman 2013). In sentiment representation, the categorical approach represents emotional states as several discrete classes such as binary (i.e., positive and negative) or as multiple categories such as Ekman's (1992) six basic emotions (anger, happiness, fear, sadness, disgust, and surprise). Classification algorithms can then be used to identify sentiment categories from texts.

The dimensional approach represents emotional states as continuous numerical values in multiple dimensions such as the valence-arousal (VA) space (Russell, 1980). The dimension of valence refers to the degree of positive and negative sentiment, whereas the dimension of arousal refers to the degree of calm and excitement. Both dimensions range from 1 (highly negative or calm) to 9 (highly positive or excited) based on the self-assessment manikin (SAM) annotation scheme (Bradley et al. 1994). For example, the following passage consisting of three sentences is associated with a valence-arousal rating of (2.5, 7.8), which displays a high degree of negativity and arousal.

(r1) *A few days ago I checked into a franchise hotel.*

(r2) *The front desk service was terrible, and they didn't know much about local attractions.*

(r3) *I would not recommend this hotel to a friend.*

Such high-arousal negative (or high-arousal positive) texts are usually of interest and could prioritized in product review systems. Dimensional sentiment analysis can accomplish this by recognizing the VA ratings of texts and rank them accordingly, thus providing more intelligent and fine-grained sentiment applications.

Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pages 225–230,
Berlin, Germany, August 7-12, 2016.

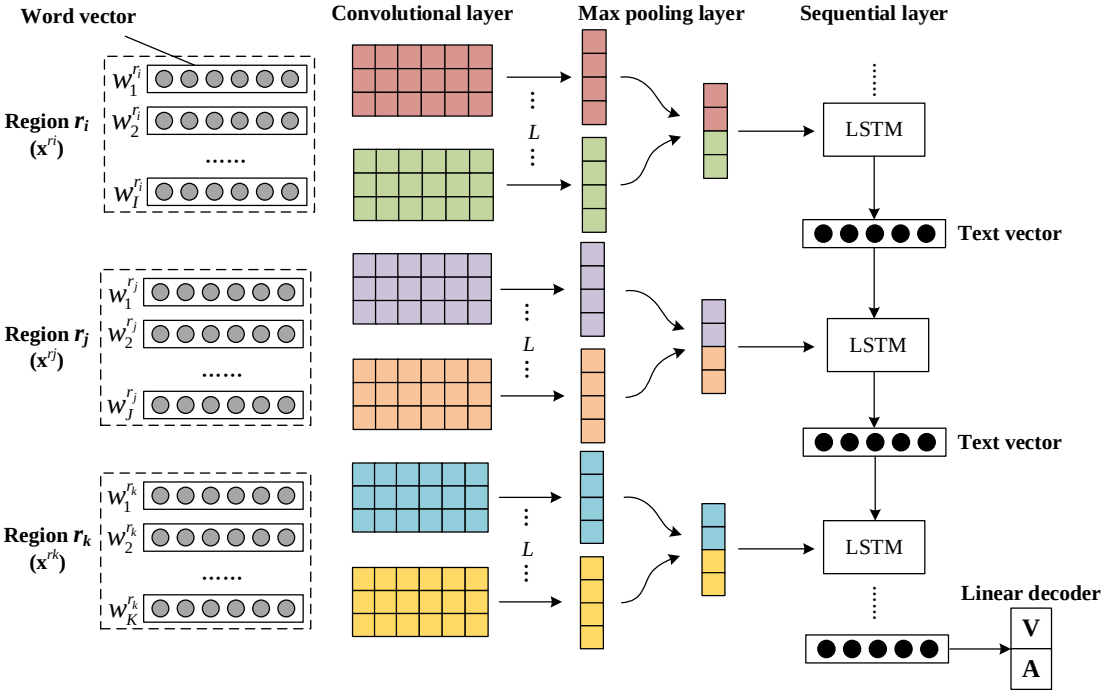

Figure 1: System architecture of the proposed regional CNN-LSTM model.

Research on dimensional sentiment analysis has addressed VA recognition at both the word-level (Wei et al., 2011; Malandrakis et al., 2011; Yu et al., 2015) and the sentence-level (Paltoglou et al., 2013; Malandrakis et al., 2013). At the word-level, Wei et al. (2011) used linear regression to transfer VA ratings from English affective words to Chinese words. Malandrakis et al. (2011) used a kernel function to combine the similarity between words for VA prediction. Yu et al. (2015) used a weighted graph model to iteratively determine the VA ratings of affective words. At the sentence level, Paltoglou et al. (2013) adopted a lexicon-based method to calculate the VA ratings of texts by averaging the VA ratings of affective words in the texts using a weighted arithmetic/geometric mean. Malandrakis et al. (2013) proposed a regression method that extracted n-gram with affective ratings as features to predict VA values for texts.

Recently, word embedding (Mikolov et al., 2013a; Mikolov et al., 2013b) and deep neural networks (NN) such as convolutional neural networks (CNN) (Kim, 2014; Kalchbrenner et al., 2014), recurrent neural networks (RNN) (Graves, 2012; Irsoy and Cardie, 2014) and long short-term memory (LSTM) (Wang et al., 2015; Liu et al., 2015) have been successfully employed for categorical sentiment analysis. In general, CNN is capable of extracting local information but may fail to capture long-distance dependency. LSTM can address this limitation by sequentially modeling texts across sentences. Such NN-based and word embedding methods have not been well explored for dimensional sentiment analysis.

This study proposes a regional CNN-LSTM model consisting of two parts, regional CNN and LSTM, to predict the VA ratings of texts. We first construct word vectors for vocabulary words using word embedding. The regional CNN is then used to build text vectors for the given texts being predicted based on the word vectors. Unlike a conventional CNN which considers a whole text as input, the proposed regional CNN uses individual sentences as regions, dividing an input text into several regions such that the useful affective information in different regions can be extracted and weighted according to their contribution to the VA prediction. For example, in the aforementioned example text, it would be useful for the system to emphasize the two sentences/regions (r2) and (r3) containing negative affective information. Finally, such regional information is sequentially integrated across regions using LSTM for VA prediction. By combining the regional CNN and LSTM, both local

(regional) information within sentences and long-distance dependency across sentences can be considered in the prediction process.

The rest of this paper is organized as follows. Section 2 describes the proposed regional CNN-LSTM model. Section 3 reports the evaluation results of the proposed method against lexicon-based, regression-based, and NN-based methods. Conclusions are finally drawn in Section 4.

2 Regional CNN-LSTM Model

Figure 1 shows the overall framework of the proposed regional CNN-LSTM model. First, the word vectors of vocabulary words are trained from a large corpus using the word2vec toolkit. For each given text, the regional CNN model uses a sentence as a region to divide the given text into R regions, i.e. $r_1,\ldots, r_i, r_j, r_k,\ldots, r_R$. In each region, useful affective features can be extracted once the word vectors sequentially pass through a convolutional layer and max pooling layer. Such local (regional) features are then sequentially integrated across regions using LSTM to build a text vector for VA prediction.

2.1 Convolutional Layer

In each region, a convolutional layer is first used to extract local n-gram features. All word embeddings are stacked in a region matrix $M \in \mathbb{R}^{d\times|V|}$, where $|V|$ is the vocabulary size of a region, and d is the dimensionality of word vectors. For example, in Fig.1, the word vectors in the regions $r_i=\{w_1^{ri}, w_2^{ri},\ldots,w_I^{ri}\}$, $r_j=\{w_1^{rj},w_2^{rj},\ldots, w_J^{rj}\}$ and $r_k=\{w_1^{rk},w_2^{rk},\ldots, w_K^{rk}\}$ are combined to form the region matrices $\mathbf{x}^{ri}$, $\mathbf{x}^{rj}$, and $\mathbf{x}^{rk}$. In each region, we use L convolutional filters to learn local n-gram features. In a window of ω words $\mathbf{x}_{n:n+\omega-1}$, a filter F_l ($1\le l\le L$) generates the feature map y_n^l as follows,

$$y_n^l = f(W^l \circ \mathbf{x}_{n:n+\omega-1} + b^l) \quad (1)$$

where $\circ$ is a convolutional operator, $W \in \mathbb{R}^{\omega\times d}$ and b respectively denote the weight matrix and bias, ω is the length of the filter, d is the dimension of the word vector, and f is the ReLU function. When a filter gradually traverses from $\mathbf{x}_{1:\omega-1}$ to $\mathbf{x}_{N+\omega-1:N}$, we get the output feature maps $\mathbf{y}^l = y_1^l, y_2^l,\ldots, y_{N-\omega+1}^l$ of filter F_l. Given varying text lengths in the regions, $\mathbf{y}^l$ may have different dimensions for different texts. Therefore, we define the maximum length of the CNN input in the corpora as the dimension N. If the input length is shorter than N, then several random vectors with a uniform distribution $U(-0.25, 0.25)$ will be appended.

2.2 Max-pooling Layer

Max-pooling subsamples the output of the convolutional layer. The most common way to do pooling it to apply a max operation to the result of each filter. There are two reasons to use a max-pooling layer here. First, by eliminating non-maximal values, it reduces computation for upper layers. Second, it can extract the local dependency within different regions to keep the most salient information. The obtained region vectors are then fed to a sequential layer.

2.3 Sequential Layer

To capture long-distance dependency across regions, the sequential layer sequentially integrates each region vector into a text vector. Due to the problem of gradients vanishing or exploding in RNN (Bengio et al., 1994), LSTM is introduced in the sequential layer for vector composition. After the LSTM memory cell sequentially traverses through all regions, the last hidden state of the sequential layer is regarded as the text representation for VA prediction.

2.4 Linear Decoder

Since the values in both the valence and arousal dimensions are continuous, the VA prediction task requires a regression. Instead of using a softmax classifier, a linear activation function (also known as a linear decoder) is used in the output layer, defined as,

$$y = W_d \mathbf{x}_t + b_d \quad (2)$$

where $\mathbf{x}_t$ is the text vector learned from the sequential layer, y is the degree of valence or arousal of the target text, and W_d and b_d respectively denote the weight and bias associated with the linear decoder.

The regional CNN-LSTM model is trained by minimizing the mean squared error between the predicted y and actual y. Given a training set of text matrix $\mathbf{X}=\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)},\ldots, \mathbf{x}^{(m)}\}$, and their VA ratings set $\mathbf{y}=\{y^{(1)}, y^{(2)}, \ldots, y^{(m)}\}$, the loss function is defined as

$$L(\mathbf{X},\mathbf{y}) = \frac{1}{2m}\sum_{i=1}^{m}\left\| h(\mathbf{x}^{(i)}) - y^{(i)} \right\|^2 \quad (3)$$

SST (English)			
Valence	RMSE	MAE	*r*
Lexicon-*w*AM	2.018	1.709	0.350
Lexicon-*w*GM	1.985	1.692	0.385
Regression-AVR	1.856	1.542	0.455
Regression-MVR	1.868	1.551	0.448
CNN	1.489	1.184	0.706
RNN	1.976	1.715	0.401
LSTM	1.444	1.151	0.717
Regional CNN-LSTM	**1.341***	**0.987***	**0.778***

***** Regional CNN-LSTM vs LSTM significantly different ($p<0.05$)

Table 1: Comparative results of different methods in SST.

CVAT (Chinese)			
Valence	RMSE	MAE	*r*
Lexicon - *w*AM	1.884	1.632	0.406
Lexicon - *w*GM	1.843	1.597	0.418
Regression-AVR	1.685	1.374	0.476
Regression-MVR	1.697	1.392	0.468
CNN	1.093	0.880	0.645
RNN	1.424	1.262	0.493
LSTM	1.135	0.939	0.641
Regional CNN-LSTM	**1.026***	**0.842***	**0.781***
Arousal	RMSE	MAE	*r*
Lexicon-*w*AM	1.232	0.985	0.268
Lexicon-*w*GM	1.243	0.996	0.263
Regression-AVR	1.154	0.862	0.286
Regression-MVR	1.128	0.842	0.289
CNN	0.991	0.788	0.453
RNN	1.024	0.816	0.290
LSTM	0.945	0.751	0.472
Regional CNN-LSTM	**0.874***	**0.689***	**0.557***

* Regional CNN-LSTM vs LSTM significantly different ($p<0.05$)

Table 2. Comparative results of different methods in CVAT.

In the training phase, a back propagation (BP) algorithm with stochastic gradient descent (SGD) is used to learn model parameters. Details of the BP algorithm can be found in (LeCun et al., 2012).

3 Experiments

This section evaluates the performance of the proposed regional CNN-LSTM model against lexicon-based, regression-based, and NN-based methods.

Datasets. This experiment used two affective corpora. i) Stanford Sentiment Treebank (SST) (Socher et al., 2013) contains 8,544 training texts, 2,210 test texts, and 1,101 validation texts. Each text was rated with a single dimension (valence) in the range of (0, 1). ii) Chinese Valence-Arousal Texts (CVAT) (Yu et al., 2016) consists of 2,009 texts collected from social forums, manually rated with both valence and arousal dimensions in the range of (1, 9) using the SAM annotation scheme (Bradley et al. 1994). The word vectors for English and Chinese were respectively trained using the Google News and Chinese wiki dumps (zhwiki) datasets. The dimensionality for both word vectors are 300.

Experimental Settings. Two lexicon-based methods were used for comparison: weighted arithmetic mean (*w*AM) and weighted geometric mean (*w*GM) (Paltoglou et al., 2013), along with two regression-based methods: average values regression (AVR) and maximum values regression (MVR) (Malandrakis et al., 2013). The valence ratings of English and Chinese words were respectively taken from the Extended ANEW (Warriner et al., 2013) and Chinese Valence-Arousal Words (CVAW) lexicons (Yu et al., 2016). A conventional CNN, RNN and LSTM were also implemented for comparison.

Metrics. Performance was evaluated using the root mean square error (RMSE), mean absolute error (MAE), and Pearson correlation coefficient (*r*), defined as

- *Root mean square error* (RMSE)

$$RMSE = \sqrt{\sum_{i=1}^{n}(A_i - P_i)^2 \Big/ n} \tag{4}$$

- *Mean absolute error* (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|A_i - P_i| \tag{5}$$

- *Pearson correlation coefficient (r)*

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{A_i - \overline{A}}{\sigma_A}\right)\left(\frac{P_i - \overline{P}}{\sigma_P}\right) \tag{6}$$

where A_i is the actual value, P_i is the predicted value, n is the number of test samples, $\overline{A}$ and $\overline{P}$ respectively denote the arithmetic mean of A and P, and σ is the standard deviation. A lower RMSE or MAE and a higher r value indicates better prediction performance. A *t-test* was used to determine whether the performance difference was statistically significant.

Comparative Results. Tables 1 and 2 respectively present the comparative results of the regional CNN-LSTM against several methods for VA prediction of texts in both English and Chinese corpora. For the lexicon-based methods, *w*GM outperformed *w*AM, which is consistent with the results presented in (Paltoglou et al., 2013). Instead of using the VA ratings of words to directly measure those of texts, the regression-based methods learned the correlations between the VA ratings of words and texts, thus yielding better performance. Once the word embedding and deep learning techniques were introduced, the performance of NN-based methods (except RNN) jumped dramatically. In addition, the proposed regional CNN-LSTM outperformed the other NN-based methods, indicating the effectiveness of sequentially integrating the regional information across regions. Another observation is that the Pearson correlation coefficient of prediction in arousal is lower than that for the valence prediction, indicating that arousal is more difficult to predict.

4 Conclusion

This study presents a regional CNN-LSTM model to predict the VA ratings of texts. By capturing both local (regional) information within sentences and long-distance dependency across sentences, the proposed method outperformed regression- and conventional NN-based methods presented in previous studies. Future work will focus on the use of a parser to identify regions so that the structural information can be further incorporated to improve the prediction performance.

Acknowledgments

This work was supported by the Ministry of Science and Technology, Taiwan, ROC, under Grant No. MOST 102-2221-E-155-029-MY3 and MOST 104-3315-E-155-002. The authors would like to thank the anonymous reviewers and the area chairs for their constructive comments.

References

Yoshua Bengio, Patrice Simard, and Paolo Frasconi. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Networks*, 5(21):57-166.

Margaret M. Bradley, and Peter J. Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25 (1): 49-59.

Rafael A. Calvo and Sidney D'Mello. 2010. Affect detection: An interdisciplinary re-view 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.

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

Alex Graves. 2012. Supervised sequence labelling with recurrent neural networks. Vol. 385, Springer.

Ozan Irsoy and Claire Cardie. 2014. Opinion mining with deep recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP-14)*, pages 720-728.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods on Natural Language Processing (EMNLP-14)*, pages 1746-1751.

Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL-14)*, pages 655-665.

Yann LeCun, Leon Bottou, Genevieve B. Orr and Klaus-Robert Muller. 2012. Efficient backprop. Neural networks: Tricks of the trade. Springer Berlin Heidelberg, 2012. 9-48.

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

Pengfei Liu, Shafiq Joty and Helen Meng. 2015. Fine-grained opinion mining with recurrent neural networks and word embeddings. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP-15)*, pages 1433-1443.

Nikos Malandrakis, Alexandros Potamianos, Elias Iosif, Shrikanth Narayanan. 2011. Kernel models for affective lexicon creation. In *Proceedings of INTERSPEECH*, pages 2977-2980.

Nikos Malandrakis, Alexandros Potamianos, Elias Iosif, Shrikanth Narayanan. 2013. Distributional semantic models for affective text analysis. *IEEE Trans. Audio, Speech, and Language Processing*, 21(11): 2379-2392.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. In *Proceedings of International Conference on Learning Representations (ICLR-13): Workshop Track.*

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

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

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.

James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161.

Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Empirical Methods on Natural Language Processing (EMNLP-13)*, pages 1631-1642.

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods*, 45(4): 1191-1207.

Xin Wang, Yuanchao Liu, Chengjie Sun, Baoxun Wang and Xiaolong Wang. 2015. Predicting polarities of tweets by composing word embeddings with long short-term memory. In *Proceedings of the 53th Annual Meeting of the Association for Computational Linguistics (ACL-15)*, pages 1343-1353.

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

Liang-Chih Yu, Jin Wang, K. Robert Lai, and Xuejie Zhang. 2015. Predicting valence-arousal ratings of words using a weighted graph method. In *Proceedings of the 53th Annual Meeting of the Association for Computational Linguistics (ACL-15)*, pages 788-793.

Liang-Chih Yu, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai and Xuejie Zhang. 2016. Building Chinese Affective Resources in Valence-Arousal Dimensions. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL/HLT-16)*.