Journal of Machine Learning Research 1:45–57, 2017 First Internal Workshop on Affective Computing (AC)

Affective State Prediction of Contextualized Concepts

Minglei Li CSMLI@COMP.POLYU.EDU.HK
Qin Lu CSLUQIN@COMP.POLYU.EDU.HK
Yunfei Long CSYLONG@COMP.POLYU.EDU.HK
Computing Department, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Lin Gui GUILIN.NLP@GMAIL.COM
College of Mathematics and Computer Science, Fuzhou University, Fuzhou, Fujian, China

Editor: Neil Lawrence and Mark Reid

Abstract

Most studies on affective analysis of text focus on the sentiment or emotion expressed by a whole sentence or document. In this paper, we propose a novel approach to predict the affective states of a described event through the predictions of the corresponding subject, action and object involved in the described event. Rather than using a sentiment label or discrete emotion categories, the affective state is represented using the three dimensional evaluation-potency-activity (EPA) model. The main idea is to use automatically obtained word embedding as word representation and to use the Long Short-Term Memory (LSTM) network as the prediction model. Compared to the linear model used in the Affective Control Theory which uses manually annotated EPA lexicon, our proposed LSTM learning method using word embedding outperforms the linear model and word embedding also performs better than EPA lexicon. Most importantly, our work shows that automatically obtained word embedding outperforms manually constructed affective lexicons.

Keywords: Affective prediction, Affective Control Theory, Word Embedding, Long short-term memory, context

1. Introduction

Affective computing aims to enable machines to recognize and express affective state like humans, a key element in artificial intelligence (Minsky, 2007). Previous studies on affective computing of text mainly focus on sentiment analysis which aims to predict either the sentiment polarity of text (Pang and Lee, 2008; Mohammad et al., 2013; Cambria et al., 2014; Alhothali and Hoey, 2015) or sentiment polarity towards a target or aspect (Wang et al., 2016). A more comprehensive study of text in terms of emotion analysis aims to classify emotions expressed in a piece of text, such as happiness, sadness, surprise, etc.(Mohammad and Turney, 2010; Calvo and Mac Kim, 2013). However, these studies focus on affective expressions in a whole piece of text, which sometimes may express the emotion of the author and is not necessarily linked to either the affective state of the subject or the object in the text. For human machine interaction, machine need to know the affective state of the subject and the object in a text that describe events (referred to as event text), such as the affective state of *"mother"* expressed in the sentence *"The mother hit the boy"* and the sentence *"The mother touched the baby"*. In this paper, we focus on affective state prediction of subjects and objects as well as the act in an event text. Subjects and objects, such as *"mother"* and *"baby"* as well as acts such as *"hit" and "touch"*, all have their generally linked concepts.

The event text is considered the context of the event. Thus, we collectively refer to this task as the affective state prediction of contextualized concepts.

The research work called the Affect Control Theory (ACT) (Heise, 2007) and related extensions (Hoey et al., 2013) do provide a good social psychological basis and a computational model to handle affective state prediction of such contextualized concepts. In the ACT, every concept is represented by the multi-dimensional Evaluation-Potency-Activity (EPA) model and every concept has a fundamental EPA representation in a specific language or culture environment. According to the ACT model, the EPA representation of a concept can change under different event context. The current EPA values under a contextualized event can be inferred through a regression model based on the fundamental EPA representation of a concept. However, one fatal drawback of the ACT theory is its lack of scalability because it requires an EPA lexicons which are manually annotated. Since EPA is a three dimensional representation in continuous scale, it is hard to obtain a large scaled EPA lexicon. Furthermore, the regression model cannot capture the complex semantic interaction of all the concepts involved in a particular context.

In this paper, we propose to use the Long Short-Term Memory (LSTM) network instead of linear regression to infer the affective state of a concept in its context. The lexicon knowledge is based on automatically obtained word embedding through unsupervised learning rather than using manually constructed EPA lexicon. Performance evaluation is conducted using an event corpus which has a subject, an act and an object annotated with EPA values as affective states. The result shows that LSTM using word embedding representation performs better than both the original ACT model and LSTM with manual EPA representation. This also indicates that there is no need to manually annotate the EPA lexicon for such kind of task based on LSTM. In fact, our experiments indicate that a seemingly sound model in theory, may not be able to work well computationally. Our proposed method can release the time-consuming manual annotation workload and achieve even better performance.

The rest of the paper is organized as follows. Section 2 introduces the related work. Section 3 presents our proposed method. Section 4 evaluates the performance of the proposed method. Section 5 concludes the paper.

2. Related Work

2.1. Affect Control Theory

ACT is a social psychological theory of human social interaction. ACT offers a rigorous methodology for modeling emotions in interaction. The models and predictions can be applied to human-computer interaction leading to the design of "socially intelligent" systems that optimize user experience and outcomes (Heise, 2007). In ACT, every concept is annotated under three dimensional evaluation-potency-activity (EPA) with the range of [-4.3,4.3].

Since the annotation is based on concept level words, the same word under different social environment may have different affective measures. For example, the concept *"dragon"* represents something good, powerful in Chinese while it represents something evil and powerful in English. Thus their corresponding EPA values may be different. ACT is also used in sociology to study the culture differences and EPA lexicons from different languages and culture environment are annotated separately and have proven to be indeed different. In general, within-cultural agreement about EPA meanings of social concepts is high even with consideration of across subgroups of society, and cultural-average EPA ratings from as little as a few dozen survey participants have been shown to be extremely stable over extended periods of time (Heise, 2010). This means that under the same language/culture environment, we can use the same EPA based lexicon.

In ACT, every event has at least three elements: subject (S), action or behavior (verb, V) and object (O). Each element is represented by a three dimensional EPA vector. For example, *"mother"* is represented as $(2.9, 1.6, 0.5)$, *"enemy"* is represented as $(-2.1, 0.8, 0.2)$ under a common culture environment, which is called the fundamental sentiment, or fundamental impression. Let us use C to denote the context of an event, and the roles of the event S, V, and O are used to denote the subject, the action (which is a transitive verb to indicate the action) and object. Thus, the fundamental impression of an event can be represented as a nine dimensional vector:

$$\boldsymbol{f_c} = [S_e, S_p, S_a, V_e, V_p, V_a, O_e, O_p, O_a], \tag{1}$$

where S_e represents the fundamental evaluation of a subject. The core assumption of ACT is that people take action to maintain their affective state towards culturally shared fundamental sentiments. In the context of *"The mother hit the boy"*, most readers would agree that the mother appears less nice (E), more powerful (P) and more active (A), which is referred to as the transient impression of the subject *"mother"*. An event can cause a transient impression, denoted by $\boldsymbol{\tau}$, based on the transient impressions of the subject, the action, and the object. The transient impression of an event C can then be expressed as:

$$\boldsymbol{\tau_c} = [S'_e, S'_p, S'_a, V'_e, V'_p, V'_a, O'_e, O'_p, O'_a] \tag{2}$$

where each element is the transient impression of corresponding subject, action and behavior. In the ACT, a feature set $\boldsymbol{t_c}$ is constructed from the fundamental sentiment of the event as:

$$\begin{aligned} \boldsymbol{t_c} = [1, S_e, S_p, S_a, V_e, V_p, V_a, O_e, O_p, O_a, S_eV_e,\\ S_eV_p, S_eV_a, S_pV_e, S_pV_p, S_pO_a, S_aV_a, V_eO_e,\\ V_eO_p, V_pO_e, V_pO_p, V_pO_a, V_aO_e, V_aO_p,\\ S_eV_eO_e, S_eV_pO_p, S_pV_pO_p, S_pV_pO_a, S_aV_aO_a] \end{aligned} \tag{3}$$

Then the ACT model obtains the transient impression $\boldsymbol{\tau_c}$ of C by a mapping function defined by

$$\boldsymbol{\tau_c} = M\boldsymbol{t_c} \tag{4}$$

where M is a parameter matrix. This is actually a linear regression model where the features of the transient impression are constructed from the fundamental impressions $\boldsymbol{f_c}$. Annotation under different languages and cultures can lead to different coefficients M. For example, the transient impression of the subject's evaluation using the US male coefficients:

$$\begin{aligned} S'_e = .98 + .48S_e - .015S_p - .015S_a + .425V_e\\ -.069V_p - .106V_a + .055O_e + \cdots \end{aligned} \tag{5}$$

The learned weights can be interpreted as how much it is affected by the corresponding dimensions. The above equation shows that the transient evaluation of the subject is mainly affected by the fundamental evaluation dimension of the subject and the evaluation dimension of the act, reflected by the positive large coefficients .48 and .425 for S_e and V_e)

2.2. Word Representation

For affective analysis, the first step is to find a proper word representation. Most of the current studies employ a sentiment lexicon for sentiment analysis and an emotion lexicon for emotion analysis. In a sentiment lexicon, each word is annotated with a sentiment label (Stone et al., 1968) or sentiment intensity (Baccianella et al., 2010; Mohammad et al., 2013; Hutto and Gilbert, 2014). In an emotion lexicon, words are labeled according to the selected emotion representation model. In the discrete emotion model, such as the Big Six model (Ekman, 1993), a word is annotated with a discrete emotion category, such as happiness, sadness, and anger, etc.. In the multi-dimensional emotion model, a word is annotated with a multi-dimensional vector, such as the three dimensional valence-arousal-dominance (VAD) values in continuous space (Bradley and Lang, 1999; Warriner et al., 2013), the three dimensional evaluation-potency-activity (EPA) values in continuous space (Heise, 1987). Compared with multi-dimensional lexicons, sentiment lexicons and discrete emotion lexicons can be obtained much more easily through manual annotation (Stone et al., 1968; Xu et al., 2008; Hastings et al., 2011), crowdsourcing (Hutto and Gilbert, 2014; Mohammad and Turney, 2013), and can be extended automatically based on some seed words (Wright et al., 2010; Tang et al., 2014; Hamilton et al., 2016; Mohammad, 2012). Multi-dimensional emotion lexicons can also be obtained through manual annotation (Bradley and Lang, 1999; Heise, 2007; Yu et al., 2016) or crowdsourcing (Warriner et al., 2013). In practice, however, the knowledge is much more difficult to acquire. It is not a simple cost issue. The multi-dimensional models are conceptually more difficult to apprehend, and the values are continuous. Thus, obtaining a high quality lexicon is much more difficult and they also face a much more severe scalability issue.

Word embedding, which represents a word by a low dimensional vector, has shown a superior advantage in encoding lexical semantic information (Mikolov et al., 2013; Pennington et al., 2014). Semantically related words are more closely located while semantic unrelated words are further apart in the vector space. For example, using word embedding, we can make use of various approximations such as: $vec(king) - vec(queen) = vec(man) - vec(woman)$ (Mikolov et al., 2013). Various models have been proposed to learn dense word vectors. All are based on the distributional hypothesis that words occur in similar context tends to have similar meanings (Harris, 1954). The various models can be divided mainly into count-based approaches and prediction-based approaches (Baroni et al., 2014). A count-based method constructs a word-context co-occurrence statistic matrix and then perform matrix factorization to obtain the final word embedding. Used features include point-wise mutual information (PMI), positive point-wise mutual information (PPMI), and log of co-occurrences, etc.. Based on matrix factorization, various algorithms have been proposed, such as decomposition of the matrix into two low dimensional matrices(Pennington et al., 2014), Singular Value Decomposition (SVD) (Levy et al., 2016), probabilistic matrix and tensor factorization (Zhang et al., 2014), and low rank approximation (Li et al., 2015). Prediction-based methods directly predict the context given the target word by maximizing the conditional probability of context words given the target or vice versa (Mikolov et al., 2013). Compared with manually constructed lexicons, the advantage of word embedding is that word embeddings are obtained automatically and the needed resources are widely available.

2.3. Composition Model

Based on the word representation model, different composition models are proposed to infer the representation of larger text units. Based on sentiments and discrete emotion lexicons, the most widely used method is to construct some features for larger text units based on heuristic rules, such as lexicon word occurrence frequencies, total sentiment score, word count under each label category (Mohammad et al., 2013). Based on vector based representations, the simplest composition method either use vector addition or weighted addition of word vectors (Mitchell and Lapata, 2010; Chen et al., 2014; Yu et al., 2016). More complex models include deep learning models, such as recursive neural network (Socher et al., 2013), recurrent neural network (Irsoy and Cardie, 2014) and convolutional neural network (Kim, 2014). Once the representation of the appropriate units is obtained, classification or regression models can be performed. In deep learning models, pre-trained word embedding are used as the input to the model.

3. Methodology

Given a word sequence consists of three parts for an event C: subject-verb-object (SVO), our objective is to predict the affective state of the subject, the verb and the object in this event. The affective state can either be sentiment, emotion category or multi-dimensional emotional representations such as VAD and EPA.

We propose to use the Long Short-Term Memory (LSTM) network (Hochreiter and Schmidhuber, 1997) for contextualized affective state prediction using word embedding as word representations. As a deep learning model, LSTM is an extension of recurrent neural network (RNN) in deep learning, which has been widely used in various NLP tasks, such as machine translation (Sutskever et al., 2014), language modeling (Graves, 2013), and sentiment analysis (Tang et al., 2015), etc.. Our proposed framework is shown in **Figure 1** using the example sentence *"mother hit boy"*. The frame consists of three layers: the input word representation layer, the LSTM layer and the output affective state layer. The input is a word vector representation $\boldsymbol{x}_t$ ($t \in [1, 2, 3]$), which can either be word embedding or a multi-dimensional word affective vector. Each LSTM cell takes the current word representation $\boldsymbol{x}_t$ and the previous output $\boldsymbol{h}_{t-1}$ of the LSTM cell as the input and outputs a hidden representation $\boldsymbol{h}_t$. The final classification or regression is performed on the last hidden representation (which is $\boldsymbol{h}_3$ in our case) based on the output type y.

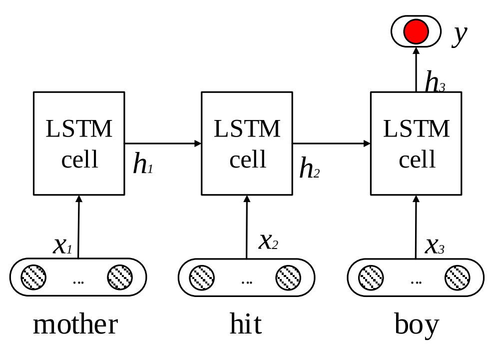


Figure 1: The LSTM model for contextualized affective prediction.

LSTM is a recurrent network very good at remembering values for either long or short durations of time. A LSTM cell at position t consists of four parts: an input gate vector $\boldsymbol{i}_t$, a forget gate vector $\boldsymbol{f}_t$, an output gate vector $\boldsymbol{o}_t$, and a cell state vector $\boldsymbol{c}_t$. Furthermore, the output of each LSTM cell is defined by the output vector $\boldsymbol{h}_t$. These vectors are defined as:

$$\boldsymbol{i}_t = \sigma(U_i \boldsymbol{x}_t + W_i \boldsymbol{h}_{t-1} + \boldsymbol{b}_i), \tag{6}$$

$$\boldsymbol{f}_t = \sigma(U_f \boldsymbol{x}_t + W_f \boldsymbol{h}_{t-1} + \boldsymbol{b}_f), \tag{7}$$

$$\boldsymbol{o}_t = \sigma(U_o \boldsymbol{x}_t + W_o \boldsymbol{h}_{t-1} + \boldsymbol{b}_o), \tag{8}$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \circ \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \circ tanh(U_c \boldsymbol{x}_t + W_c \boldsymbol{h}_{t-1} + \boldsymbol{b}_c), \tag{9}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \circ tanh(\boldsymbol{c}_t). \tag{10}$$

where σ is the sigmoid activation function, $\circ$ denotes Hadamard product. $\boldsymbol{b}_i, \boldsymbol{b}_f, \boldsymbol{b}_o, \boldsymbol{b}_c$ are the bias. Based on the current input $\boldsymbol{x}_t$ and the previous cell's output $\boldsymbol{h}_{t-1}$, it computes the input gate vector $\boldsymbol{i}_t$, forget gate vector $\boldsymbol{f}_t$, output state vector $\boldsymbol{o}_t$. Current output $\boldsymbol{h}_t$ is computed using **Formula 10**. $U_i, U_f, U_o, U_c, W_i, W_f, W_o, W_c$ are the model matrix parameters that are learned during training the model. A regression or classification is performed on the final LSTM cell and outputs the predicted affective state y. For each affective state of each role in an event, we train a LSTM model. The stochastic gradient descent method is used to train the model and the code is implemented based on Keras[1].

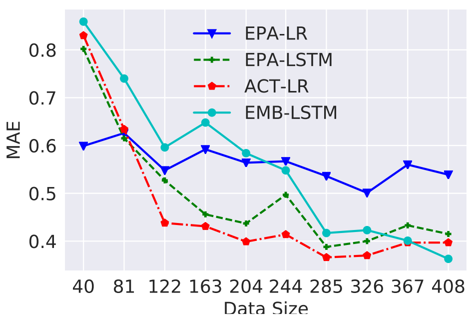
F	Model	**Subject**			**Behavior**			**Object**		
		E	P	A	E	P	A	E	P	A
VAD	LR	.682(.059)	.728(.031)	.539(.023)	.601(.093)	.630(.053)	.496(.041)	.478	.791	.658
VAD	LSTM	1.078(.039)	.751(.031)	.692(.032)	1.047(.229)	.637(.063)	.710(.055)	.561	.913	.754
EPA	LR	.556(.051)	.326(.011)	.309(.006)	.467(.048)	.263(.014)	.256(.009)	.301	.357	.349
EPA	LSTM	.485(.045)	.372(.011)	.341(.029)	.410(.044)	.322(.027)	.289(.011)	.282	.352	.318
ACT	LR	.385(.030)	**.325(.020)**	.313(.010)	**.315(.033)**	.267(.014)	**.257(.009)**	.263	.355	.360
EMB	LSTM	**.363(.036)**	.353(.015)	**.274(.009)**	.348(.035)	**.241(.024)**	.261(.041)	**.255**	**.277**	**.265**

Table 1: Affective prediction of contextualized concepts based on different features and composition functions. The evaluation metric is **MAE**.The number in the parenthesis is the standard deviation of the MAE of the five runs.For better display, we do not include the standard deviation for the result of Object.

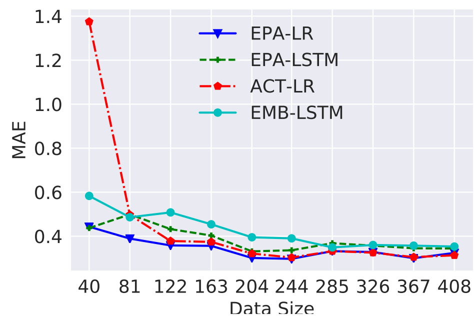
4. Performance Evaluation

Evaluation of our proposed method is conducted using the ACT corpus from (Heise, 2007) as the golden answer. In this corpus, every sample forms a 3-tuple consisting of three words which describes an event in the form of SVO, such as *"vampire enslave heroine"*. The total size of the corpus is 515 sentences in SVO form with a vocabulary size of 106. Every word under an event is annotated with the EPA values as $\boldsymbol{\tau}$, the transient impressions. The annotation was conducted by about 25 females and 25 males of Americans using the semantic differential scheme. We take the average of the annotations from both male and female annotators as the final representation. The fundamental EPA values of the 106 words are also provided by (Heise, 2007). This can be used to

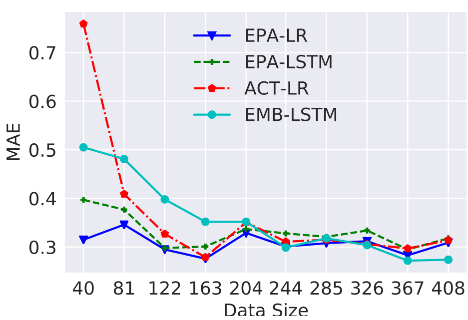
1. https://keras.io



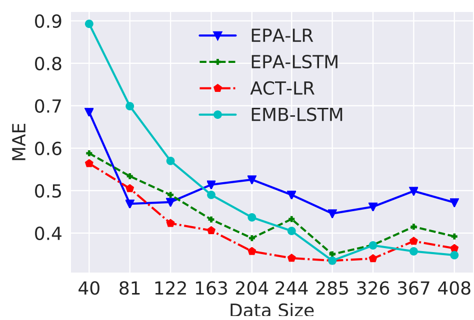
(*a*) E dimension of subject



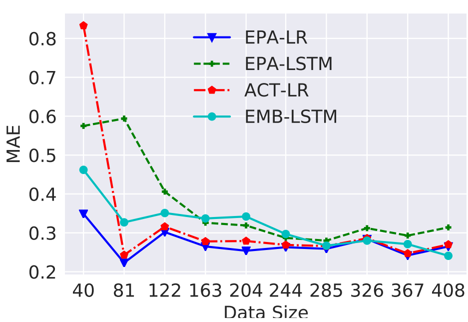
(*b*) P dimension of subject



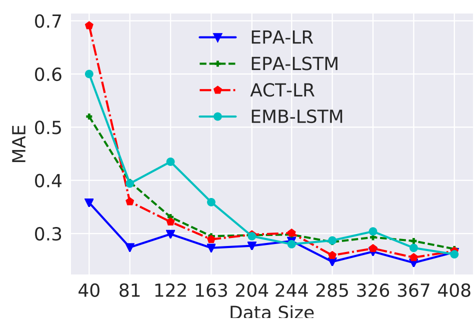
(*c*) A dimension of subject



(*d*) E dimension of behavior



(*e*) P dimension of behavior



(*f*) A dimension of behavior

Figure 2: The performance on different affective dimensions of Subject and Behavior when varying the training data size.

train models such as Formula 4. Since EPA are scalar values, we perform regression on the output of the last LSTM cell.

For word embedding, we use the available Glove 840B word embedding which is trained on a corpus of 840 billion tokens based on matrix factorization (Pennington et al., 2014). The embedding dimension is 300, denoted as emb300d. Note that only 99 out of the 106 words in the ACT corpus appear in Glove 840B collection. To focus on the effectiveness of representation and eliminate the effect of coverage problem, we only use the overlap vocabulary set of the embedding and the ACT corpus. So, the final evaluation corpus size is actually 408 sentences.

Our evaluation is to check the performance of the predicted transient impressions (the EPA values) of the subject, verb and object in each sample sentence. We perform 5-fold cross validation and use the best parameters obtained through manually tuning. The evaluation metric is the Mean Absolute Error (MAE).

Our proposed method, denoted as EMB-LSTM, is compared with the following five baseline methods:

1. VAD-LR: This is a linear regression based method using the VAD affective model. The features include the concatenation of the VAD values of S, V, and O. The VAD lexicon is from (Warriner et al., 2013) which includes about 13K words annotated in the three dimensions of valence-arousal-dominance.

2. VAD-LSTM: This is LSTM based method using the VAD lexicon as word representation. The VAD data is the same as that of VAD-LR.

3. EPA-LR: This is a linear regression based method using the EPA affective model as word representation. The EPA lexicon is provided by (Heise, 2007). The features use the concatenation of the EPA values of S, V, and O.

4. EPA-LSTM: This is a LSTM based method using EPA as word representation.

5. ACT-LR: This is a linear regression based method. The input features are $\boldsymbol{t}$ defined in **Equation 3** constructed from the EPA values.

The evaluation result is shown in **Table 1**. The value in the parenthesis is the standard deviation of the MAE in the five runs, which can indicate the robustness of different models. We use this to test whether overfitting occurs in LSTM model because our training data size is too small and the parameter size of LSTM plus word embedding is quite large. If the standard deviation is too large, overfitting may occur. Comparing between the standard deviations of different models indicate taht the standard deviations of LSTM are similar with the other models. We can conclude that no overfitting occurs in the LSTM model. Comparing between the manual VAD and EPA representations, EPA performs much better than VAD under both the linear regression model and the LSTM model. This is expected because the predicted affective state is represented by the EPA model in which the training data is more relevant. This, however, may also be because the VAD lexicon data is obtained through crowdsourcing, which has lower quality than the EPA lexicon. The ACT features perform better than EPA features on the evaluation (E) dimension while they perform comparatively on potency (P) and activity (A). The difference between ACT and EPA is only in the additional features on the interaction of the EPA values in ACT. Thus, ACT has better performance than that of EPA based method is also understandable.

Overall, our proposed EMB-LSTM has the best performance in 6 columns out of 9 as shown in bold in **Table 1**. Overall EMB-LSTM performs better with large margins compared to the second performer ACT-LR. ACT-LR using regression performs slightly better on two columns. One is the P dimension of the subject and the other is the E dimension of the object. Comparing between EPA-LSTM and EMB-LSTM, EMB-LSTM has much better performance. This clearly indicates the advantage of using word embedding than using the manually annotated lexicon. This analysis validates the effectiveness of the proposed model of using both LSTM and word embedding. Most importantly, our experimental result indicates that automatically obtained word embedding outperforms manually annotated EPA and VAD lexicon. It should be pointed out that this experiment does not consider the coverage issue because of the limited size of the ACT corpus. If coverage issue is considered, the advantage of word embedding should be even more obvious.

Machine learning methods normally require certain amount of training data to learn sufficient information. In the next experiment, we further examine the effect of training data size to the performance of different models. We perform the experiment on the same ACT corpus by varying the training data size starting from 40 sentences to the whole dataset of 408 sentences and run 5-fold cross validation for each affective dimension of subject and behavior. Result in **Figure 2** shows that as the training data size increases, the performance of all the models improves. However, as the data size increases, EMB-LSTM shows its better learning ability as its performance continue to improve. Overall, when the sample number reaches about 285 to 300, EMB-LSTM performs the best on most dimensions. Due to the limitation of our dataset size, we cannot validate the performance on larger data size. But, the trend of performance in this experiment clearly indicates that our LSTM with word embedding performs the best and more training data will likely to show even more advantage of EMB-LSTM. On the other hand, the linear regression model using EPA has very steady performance when varying the training data size. This also suggests that regression model can be useful if training data size is small.

5. Conclusion

In this paper, we propose a novel approach to predict the affective states of the subject, action and object involved in a described event. The affective states are represented using three-dimensional EPA model instead of using discrete sentiment labels or emotion labels. The main idea of our proposed approach is to use automatically obtained word embedding as word representation and to use the Long Short-Term Memory network as the prediction model. Compared to the linear model used in ACT which uses a manually annotated EPA lexicon, our proposed LSTM with word embedding outperforms the linear model on most affective dimensions. Most importantly, our work indicates that automatically obtained word embedding outperforms manually constructed affective lexicons. In fact, our experiments indicate that a seemingly sound model in theory, may not be able to work well computationally. Our study shows that there no need to manually annotate EPA lexicons as manual annotation of data in three-dimensional continuous space is extremely time-consuming, hard to maintain quality, and difficult to scale up.

One limitation of current work is that the proposed model can only handle structured sequence with the form of subject-behavior-object. Additional information in sentences such as time, place and other descriptive features are not being used. Extending our model to more complex event descriptions will be our future direction.

Acknowledgments

Acknowledgements go here. This work is supported by HK Polytechnic University (PolyU RTVU and GRF PolyU 15211/14E).

References

Areej Alhothali and Jesse Hoey. Good News or Bad News: Using Affect Control Theory to Analyze Readers' Reaction Towards News Articles. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*, pages 1548–1558, Denver, Colorado, USA, 2015.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, volume 10, pages 2200–2204, 2010.

Marco Baroni, Georgiana Dinu, and Germn Kruszewski. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 238–247, Baltimore, MD, USA, 2014.

Margaret M Bradley and Peter J Lang. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical report, Technical report C-1, the center for research in psychophysiology, University of Florida, 1999.

Rafael A. Calvo and Sunghwan Mac Kim. Emotions in text: dimensional and categorical models. *Computational Intelligence*, 29(3):527–543, 2013.

Erik Cambria, Daniel Olsher, and Dheeraj Rajagopal. SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. 2014.

Tao Chen, Ruifeng Xu, Qin Lu, Bin Liu, Jun Xu, Lin Yao, and Zhenyu He. A Sentence Vector Based Over-Sampling Method for Imbalanced Emotion Classification. In *Computational Linguistics and Intelligent Text Processing*, pages 62–72. Springer, 2014. ISBN 3642549020.

Paul Ekman. Facial expression and emotion. *American psychologist*, 48(4):384, 1993. ISSN 1935-990X.

Alex Graves. Generating Sequences With Recurrent Neural Networks. *arXiv:1308.0850 [cs]*, August 2013. 00069.

William L. Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky. Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 595–605, Austin, Texas, USA, 2016.

Zellig S. Harris. Distributional structure. *Word*, 1954.

Janna Hastings, Werner Ceusters, Barry Smith, and Kevin Mulligan. The emotion ontology: enabling interdisciplinary research in the affective sciences. In *Modeling and Using Context*, pages 119–123. Springer, 2011. ISBN 3642242782.

David R. Heise. Affect control theory: Concepts and model. *The Journal of Mathematical Sociology*, 13(1-2):1–33, 1987. ISSN 0022-250X. doi: 10.1080/0022250X.1987.9990025.

David R. Heise. *Expressive Order: Confirming Sentiments in Social Actions*. Springer US, Boston, MA, 2007. ISBN 978-0-387-38177-0. 00187.

David R. Heise. *Surveying Cultures: Discovering Shared Conceptions and Sentiments*. John Wiley & Sons, March 2010. ISBN 978-0-470-57576-5. 00081.

Sepp Hochreiter and Jrgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997. 01063.

Jesse Hoey, Tim Schroder, and Areej Alhothali. Bayesian affect control theory. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 166–172. IEEE, 2013. 00020.

C. J. Hutto and Eric Gilbert. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM)*, 2014.

Ozan Irsoy and Claire Cardie. Opinion Mining with Deep Recurrent Neural Networks. pages 720–728, 2014.

Yoon Kim. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1746–1751, 2014.

Omer Levy, Anders Sgaard, and Yoav Goldberg. Reconsidering Cross-lingual Word Embeddings. August 2016.

Shaohua Li, Jun Zhu, and Chunyan Miao. A Generative Word Embedding Model and its Low Rank Positive Semidefinite Solution. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1599–1609, 2015. doi: 10.18653/v1/D15-1183.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Proceedings of 27th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 3111–3119, Nevada, United States, 2013.

Marvin Minsky. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon & Schuster, New York, reprint edition edition, November 2007. ISBN 978-0-7432-7664-1.

Jeff Mitchell and Mirella Lapata. Composition in distributional models of semantics. *Cognitive Science*, 34(8):1388–1429, November 2010. ISSN 1551-6709. doi: 10.1111/j.1551-6709.2010.01106.x.

Saif M Mohammad. # Emotional tweets. pages 246–255. Association for Computational Linguistics, 2012.

Saif M Mohammad and Peter D Turney. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. pages 26–34. Association for Computational Linguistics, 2010.

Saif M Mohammad and Peter D Turney. Crowdsourcing a wordemotion association lexicon. *Computational Intelligence*, 29(3):436–465, 2013. ISSN 1467-8640.

Saif M Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. In *Proceedings of the 7th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT*, pages 321–327, 2013.

Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008. ISSN 1554-0669.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014.

Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, pages 1631–1642, 2013.

Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Ogilvie. The General Inquirer: A Computer Approach to Content Analysis. *Journal of Regional Science*, 8(1):113–116, 1968. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9787.1968.tb01290.x/full.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to Sequence Learning with Neural Networks. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 3104–3112, 2014.

Duyu Tang, Furu Wei, Bing Qin, Ming Zhou, and Ting Liu. Building Large-Scale Twitter-Specific Sentiment Lexicon : A Representation Learning Approach. In *Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers (COLING)*, pages 172–182, Dublin, Ireland, 2014.

Duyu Tang, Bing Qin, and Ting Liu. Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1422–1432, 2015.

Shuai Wang, Zhiyuan Chen, and Bing Liu. Mining Aspect-Specific Opinion Using a Holistic Lifelong Topic Model. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, pages 167–176, Republic and Canton of Geneva, Switzerland, 2016. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-4143-1.

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

Hyrum K. Wright, Miryung Kim, and Dewayne E. Perry. Validity concerns in software engineering research. In *Proceedings of the FSE/SDP workshop on Future of software engineering research*, pages 411–414. ACM, 2010.

Linhong Xu, Hongfei Lin, Pan Yu, Hui Ren, and Jianmei Chen. Constructing the Affective Lexicon Ontology. *Journal of the China Society for Scientific and Technical Information*, 2:006, 2008.

Liang-Chih Yu, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai, and Xuejie Zhang. Building Chinese Affective Resources in Valence-Arousal Dimensions. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 540–545, 2016.

Jingwei Zhang, Jeremy Salwen, Michael R. Glass, and Alfio Massimiliano Gliozzo. Word Semantic Representations using Bayesian Probabilistic Tensor Factorization. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1522–1531, 2014.