

deepCybErNet at EmoInt-2017: Deep Emotion Intensities in Tweets

Vinayakumar R and Premjith B and Sachin Kumar S and Soman K P

Center for Computational Engineering and Networking,
Amrita School of Engineering, Amrita Vishwa Vidyapeetham,
Amrita University, India
vinayakumarr77@gmail.com

Prabaharan Poornachandran

Center for Cyber Security Systems and Networks,
Amrita School of Engineering, Amrita Vishwa Vidyapeetham,
Amrita University, India

Abstract

This working note presents the methodology used in deepCybErNet submission to the shared task on Emotion Intensities in Tweets (EmoInt) WASSA-2017. The goal of the task is to predict a real valued score in the range [0-1] for a particular tweet with an emotion type. To do this, we used Bag-of-Words and embedding based on recurrent network architecture. We have developed two systems and experiments are conducted on the Emotion Intensity shared Task 1 data base at WASSA-2017. A system which uses word embedding based on recurrent network architecture has achieved highest 5 fold cross-validation accuracy. This has used embedding with recurrent network to extract optimal features at tweet level and logistic regression for prediction. These methods are highly language independent and experimental results shows that the proposed methods is apt for predicting a real valued score in than range [0-1] for a given tweet with its emotion type.

1 Introduction

Internet has become an essential platform to carry out daily activities to our lives. People use social media resources like Twitter, Facebook, WhatsApp, Hike, WeChat etc. to share their language such as views or emotions, stance over issues, reviews related to products, services, blogs etc. In recent days, the amount of language sharing through the internet is ubiquitous. This necessitates the need of analyzing reviews to identify the emotions including estimating the degree to which an emotion is expressed in text. Unlike natural language, the user reviews are small; rich

information is represented through nonstandard language such as emoticons, emojis, creatively spelled words (happee), and hash-tagged words (#happy). These factors can make a high influence on the social and economic behavior worldwide like real-world applications such as marketing, e-Governance, business intelligence, social analysis and applications in Natural Language Processing (NLP) - information extraction, question answering, textual entailment, etc. Many methods have been introduced by researchers for emotion annotation work. This gives binary labels for the given text (Alm et al., 2005), (Aman and Szpakowicz, 2007; Brooks et al., 2013), (Neviarouskaya et al., 2009), (Bollen et al., 2011), (Summa et al., 2016). only one annotation work exists for providing a real valued score as annotation for a given text (Strapparava and Mihalcea, 2007). This was a task included in the SemEval-2007 shared task. Many methods devised for automatic emotion classification (Werbos, 1990), (Summa et al., 2016), (Mohammad, 2012), (Bollen et al., 2011), (Aman and Szpakowicz, 2007), (Brooks et al., 2013). However, only less amount work exists on emotion regression other than SemEval-2007 shared task (Strapparava and Mihalcea, 2007).

In this paper, we use Bag-of-Words (BOW) and a Bag-of-Words (BOW) based recurrent embedding system for predicting a real valued score in the range [0-1]. In first case, BOW is used to obtain the feature representation for the tweets and classification is done using logistic regression. We also employed an RNN and LSTM based method for mining the features at tweets level. These methods are language independent. So irrespective of the language, we can use these approaches for finding the stance of micro blogging posts.

The rest of the paper is organized as follows. Section 2 discusses information of shared task. Section 3 discusses the proposed methodology.

Dataset	Anger	Fear	Joy	Sadness
Training	857	1147	823	786
Development	84	110	79	74
Testing	760	995	714	673

Table 1: Statistics of Tweet Emotion Intensity dataset

Section 4.2 provides experimental analysis and results and at last conclusion is placed in Section 5.

2 Task description

The Emotion Intensity Task is a shared task of 8th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA 2017) in conjunction with the EMNLP 2017 conference (Mohammad and Bravo-Marquez, 2017). The aim of the task is to obtain a real valued score in the range [0-1] for the given tweet including an emotion type. The tweets in training, validation and testing are from four different categories such as anger, fear, joy, sadness. Each tweet has an emotion type with its score in the range [0-1], where 0 denotes that the tweet has maximally away from its emotion and 1 denotes that the tweet has maximally closer to its emotion. The detailed statistics of the data set is described in Table 1.

3 Methodology

This section provides the information of the proposed approach for predicting a real valued score in the range [0-1] for a given tweet with an emotion type. We used two approaches; (1) Bag-of-words (BoW) based word embedding (2) Recurrent Neural Network (RNN) based word embedding

3.1 Bag-of-words based system for Emotion Intensities in Tweets

The embedding size was set to 256 so that each word is now represented using a 256 dimension vector and word length to 70. Anger, Fear, Joy and Sadness have 857, 1147, 823 and 786 instances. We constructed matrix of shape 857×70 , 1147×70 , 823×70 and 786×70 for training instances and 84×70 , 110×70 , 79×70 and 74×70 for development instances. Next, we replace each word with their corresponding word embedding and this forms an input tensor of shape $857 \times 70 \times 256$, $1147 \times 70 \times 256$, $823 \times 70 \times 256$ and $786 \times 70 \times 256$ for training instances and $84 \times 70 \times 256$, $110 \times 70 \times 256$, $79 \times 70 \times 256$ and

$74 \times 70 \times 256$ for development instances. At last, an input tensor is transformed to matrix of shape 857×256 , 1147×256 , 823×256 and 786×256 for training instances and 84×256 , 110×256 , 79×256 and 74×256 for development instances using max-pooling approach. These matrices are passed to logistic regression and a real valued score is chosen for a given tweet with an emotion type using argmax function.

3.2 Recurrent neural network (RNN) based system for Emotion Intensities in Tweets

Recurrent neural network is largely used deep learning architecture for sequence data modeling. This has achieved significant results in various tasks exists in the field of natural language processing (LeCun et al., 2015). It generally looks same as feed forward networks (FFN), additionally contains self-recurrent connection in units (Elman, 1990). This cyclic loop carries out information from one time-step to another. Consequently, RNN are able to learn the temporal patterns by considering the past information in estimating the present states. Generally, RNN takes input as $x_t \in R^n$ and $h_{t-1} \in R^m$ of arbitrary length to compute succeeding hidden state vector h_t by using the following formulae recursively.

$$h_t = f(w_{xh}x_t + w_{hh}h_{t-1} + b) \quad (1)$$

$$o_t = sf(w_{oh}h_t + b_{ot}) \quad (2)$$

Where f is the nonlinear activation function, particularly logistic sigmoid function (σ) applied on element wise, h_{i0} is usually initialized to 0 at time-step t_0 and $w_{xh} \in R^{m \times n}$, $w_{hh} \in R^{m \times m}$ and $b \in R^m$ are arguments of affine transformation. Here o is the output at time step t .

Using RNN approach, a system was implemented for predicting a real valued score in the range [0-1] for emotional intensities in tweets. By following the aforementioned mechanism, we constructed an input tensor of shape $857 \times 70 \times 256$, $1147 \times 70 \times 256$, $823 \times 70 \times 256$ and $786 \times 70 \times 256$ for training instances and $84 \times 70 \times 256$, $110 \times 70 \times 256$, $79 \times 70 \times 256$ and $74 \times 70 \times 256$ for development instances. So the embedding matrix for the tweets of size 70×256 in both training and development are now reduced to 256 dimensional vectors. So, embedding matrices of size 857×256 , 1147×256 , 823×256 and 786×256 were used as training samples and

84X256, 110X256, 79X256 and 74X256 were taken as development instances and then fed into the RNN layer followed by logistic regression for prediction.

3.3 Long short-term memory based system for Emotion Intensities in Tweets

RNN issues vanishing and exploding gradient issue in memorizing long-term dependencies (Bengio et al., 1994). To reduce, (Hochreiter and Schmidhuber, 1997) has introduced long short-term memory (LSTM). Unlike RNN simple units in recurrent hidden layer, LSTM has introduced a memory block. A memory block is a complex processing unit that contains one or more memory cell, adaptive gates such as input gate and output gate and Constant Error Carousel (CEC). A memory block stores an information and updates them across time-steps based on the input and output gates. Input and output gate controls the input and output flow of information to a memory cell. Additionally, it has a built-in value as 1 for constant Error carousel (CEC). This value will be activated when in the absence of value from the outside the signal. Moreover, (Gers et al., 1999) introduced forget gate, (Gers et al., 2002) introduced peephole connections to the memory block in LSTM. A forget gate facilitates to forget or reset the values across time steps and peephole connections helps to learn precise timing of the outputs. The newly proposed architecture has performed well in learning long-range temporal dependencies in various artificial intelligence (AI) tasks (LeCun et al., 2015). Generally, at each time step an LSTM network considers the following 3 inputs; x_t, h_{t-1}, c_{t-1} and outputs h_t, c_t through the following equations

$$i_t = \sigma(w_i x_t + U_i h_{t-1} + V_i m_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(w_f x_t + U_f h_{t-1} + V_f m_{t-1} + b_f) \quad (4)$$

$$o_t = \sigma(w_o x_t + U_o h_{t-1} + V_o m_{t-1} + b_o) \quad (5)$$

$$\tilde{m}_t = \tanh(w_m x_t + U_m h_{t-1} + b_m) \quad (6)$$

$$m_t = f_t^i \odot m_{t-1} + i_t \odot \tilde{m} \quad (7)$$

$$h_t = o_t \odot \tanh(m_t) \quad (8)$$

Where x_t is the input at time step t , σ is sigmoid non-linear activation function, \tanh is hyperbolic tangent non-linear activation function, \odot denotes element-wise multiplication. Concretely, at $t = 0$

Method	Emotion	Pearson	Spearman
Bow	Anger	0.677	0.697
	Fear	0.675	0.685
	Joy	0.601	0.621
	Sadness	0.657	0.647
RNN	Anger	0.718	0.707
	Fear	0.715	0.75
	Joy	0.601	0.721
	Sadness	0.707	0.71
LSTM	Anger	0.721	0.736
	Fear	0.72	0.753
	Joy	0.621	0.725
	Sadness	0.737	0.724

Table 2: 5-fold cross validation with embedding vector size 128

hidden and memory cell state vectors such as h_0 and c_0 are initialized to 0.

We followed subsections 3.1 and 3.2 to develop a LSTM based system for predicting a real valued score in the range [0-1] for a given emotion including its emotion type. This system is constructed by simple replacing RNN layer with LSTM.

4 Experiments

All deep learning architecture are trained using GPU enabled TensorFlow (Abadi et al., 2016) with backpropagation through time (BPTT) (Werbos, 1990).

4.1 Parameter Selection

To choose optimal parameter for embedding size, the LSTM model is trained with embedding size 128 and 256 and the performance of them is evaluated on the development data set. The detailed evaluation results are displayed in Tables 2 and 3. We didn't use any hyper parameter tuning mechanism for tweet length instead we used static length 70 in all our experiments.

4.2 Evaluation results

We have submitted one run based on LSTM based recurrent embedding system to WASSA2017 and the detailed results is displayed in Tables 4 and 5

Analysis of training results and testing results showed that there is a significant difference in the performance measure. This is due to the overfitting of the model to the training data because, a deep learning framework requires huge amount of data to learn the features. Unavailability of such

Method	Emotion	Pearson	Spearman
BoW	Anger	0.681	0.71
	Fear	0.682	0.695
	Joy	0.611	0.632
	Sadness	0.661	0.654
RNN	Anger	0.721	0.714
	Fear	0.724	0.761
	Joy	0.613	0.742
	Sadness	0.714	0.721
LSTM	Anger	0.731	0.741
	Fear	0.741	0.764
	Joy	0.634	0.732
	Sadness	0.739	0.731

Table 3: 5-fold cross validation with embedding vector size 256

sufficient training data samples caused the overfitting of the system. This in turn affected the accuracy of prediction.

5 Conclusion

This working note has presented a language independent approach based on BoW and recurrent based embedding for predicting a real valued score in the range [0-1] for a given tweet with an emotion type. LSTM network has outperformed both bag-of-words embedding and recurrent based embedding mechanism. This is primarily due to the fact that LSTM has capability to learn long-temporal dependencies across time steps. Due to less number of instances in training data, the accuracy of the proposed mechanism is less. Though, the efficacy of embedding's of RNN and LSTM is considerable and paves the manner in future to use for predicting real valued score in the range [0-1] with more training instances including its emotion type. To justify that the proposed deep learning mechanism has capability to perform better with large amount of instances will be remained as one direction towards future work.

References

Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*. Savannah, Georgia, USA.

Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, pages 579–586.

Saima Aman and Stan Szpakowicz. 2007. Identifying expressions of emotion in text. In *International Conference on Text, Speech and Dialogue*. Springer, pages 196–205.

Yoshua Bengio, Patrice Simard, and Paolo Frasconi. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks* 5(2):157–166.

Johan Bollen, Huina Mao, and Alberto Pepe. 2011. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *ICWSM* 11:450–453.

Michael Brooks, Katie Kuksenok, Megan K Torkildson, Daniel Perry, John J Robinson, Taylor J Scott, Ona Anicello, Ariana Zukowski, Paul Harris, and Cecilia R Aragon. 2013. Statistical affect detection in collaborative chat. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, pages 317–328.

Jeffrey L Elman. 1990. Finding structure in time. *Cognitive science* 14(2):179–211.

Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. 1999. Learning to forget: Continual prediction with lstm .

Felix A Gers, Nicol N Schraudolph, and Jürgen Schmidhuber. 2002. Learning precise timing with lstm recurrent networks. *Journal of machine learning research* 3(Aug):115–143.

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

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521(7553):436–444.

Saif M Mohammad. 2012. # emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*. Association for Computational Linguistics, pages 246–255.

Saif M Mohammad and Felipe Bravo-Marquez. 2017. Emotion intensities in tweets.

Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. 2009. Compositionality principle in recognition of fine-grained emotions from text. In *ICWSM*.

Emotion	Pearson	Spearman	Average Pearson	Average Spearman
Anger	0.176	0.155	0.076	0.071
Fear	0.023	0.011		
Joy	-0.019	0.008		
Sadness	0.124	0.108		

Table 4: Test results in range [0-1]

Emotion	Pearson	Spearman	Average Pearson	Average Spearman
Anger	0.19	0.164	0.14	0.134
Fear	0.077	0.061		
Joy	-0.057	0.071		
Sadness	0.235	0.238		

Table 5: Test results in range [0.5-1]

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*. Association for Computational Linguistics, pages 70–74.

Anja Summa, Bernd Resch, Geoinformatics-Z GIS, and Michael Strube. 2016. Microblog emotion classification by computing similarity in text, time, and space. In *Proceedings of the Workshop on Computational Modeling of Peoples Opinions, Personality, and Emotions in Social Media*. pages 153–162.

Paul J Werbos. 1990. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE* 78(10):1550–1560.