DEFT 2017 - Texts Search @ TALN / RECITAL 2017: Deep Analysis of Opinion and Figurative language on Tweets in French

Vinayakumar R, Sachin Kumar S, Premjith B, Prabaharan P, and Soman K P
Center for Computational Engineering and Networking,
Center for Cyber Security Systems and Networks,
Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Amrita University, India
vinayakumarr77@gmail.com

ABSTRACT

The working note discusses the description of our language independent system submitted to the DEFT 2017 three shared tasks on Opinion analysis and figurative language in twitter tweets in French. We use embedding of bag-of-words method with a family of recurrent neural networks to analysis of tweets occurred around on the analysis of opinion and figurative language. We developed three systems for each shared task and each system focuses on Opinion analysis and figurative language substantially at the tweets level only. A family of recurrent neural network extracts features in each tweet and classified them using logistic regression. On task1, our system achieved Macro f-score of 0.276, 0.228, and 0.21 with long short-term memory (LSTM) for extracting features from tweets and logistic regression for classification. On task2 our system achieved Macro f-score 0.475, 0.470, 0.476 with recurrent neural network (RNN) for extracting features from tweets and logistic regression for classification. And on task3 our system achieved Macro f-score 0.22, 0.232, 0.231 with gated recurrent unit (GRU) for extracting features from tweets and logistic regression for classification. Apart from results, this working note give valuable deep insights in to applicability of deep learning mechanisms for Sentimental analysis (SA) or Opinion mining (OM). Moreover the proposed method typically serves as a language independent method.

Mots-clés : analyse de sentiment et d'opinion, langage figuratif, français, Twitter, apprentissage profond : recurrent neural network (RNN), long short-term memory (LSTM), gated recurrent unit (GRU)

KEYWORDS: Sentimental analysis (SA) or Opinion mining (OM): opinion and Figurative language, French, Twitter, language independent, deep learning: recurrent neural network (RNN), long short-term memory (LSTM), gated recurrent unit (GRU)

1 Introduction

In today’s world internet has become a primary medium to carry out daily activities. As a result, examining the information flow in each activity has an important in associated with social services like Twitter, Facebook, LinkedIn, and chat services like Hike, Whatapp and Wechat [15]. Sentimental analysis (SA) or Opinion mining (OM) related to Tweets in Twitter is the task of studying the subjectivity and polarity of them [1]. Sentimental analysis has plays an important role in various fields like e-commerce companies can use user reviews on their cite to predict the demand, political parties may use it for knowing the citizens reviews and based on the decision taken and so many. The area has been flourished in recent days in automatic language processing due to the amplification of user-generated texts in World Wide Web (WWW) and the possibility to largely scatter emotions, opinions and evaluations [2]. Certainly, the availability of sources of user opinions in user-generated texts has been multiplying. In perspective of this plenitude of information and sources, mechanizing the synthesis of multiple opinions becomes vital to obtain a good summary of opinions on a given subject. Generally, the existing approaches have really performed well on the automated classification of the subjective or objective character of a document [2]. On the other side, the performance on polarity analysis task undertaking continues to be indecisive. Thus further enhancement in polarity analysis has required discoursing more appropriate language devices such as figurative language.Generally literal and figurative are two types of language. Literal language defines the meaning of a test explicitly. Contrary to literal language, figurative language redirects its appropriate intending to give it an importance called figurative or imagery, such as irony, sarcasm, humor, analogy, ambiguity, metaphor or punctuations. The state of humorous has generally includes irony. It can exist in social network services, stories, user reviews and many others. This has made analyzing sentiment of figurative documents is still arduous for computers as well as human beings. Towards enhancing the performance of opinion analysis systems, this has been served as an active area of research [3, 6, 8].

Irony is one of the most arduous domains extensively studied in the field of philosophy and linguistics [4]. Defining a unique satisfactory definition for irony is considered as often difficult one. The expert says irony as a figure of rhetoric by which one says the opposite of what one wants to make understand. For example, to express a negative opinion on his mobile phone, one can use a literal form ("This phone is a disaster") or a pictorial form ("What a great phone!"). In computational linguistics, irony is a generic term used to describe a set of figurative phenomena including sarcasm, even if the latter expresses itself with more bitterness and aggressiveness [5]. Detecting and characterizing the figurative language and its part in perusal of feelings has been the topic of several evaluation campaigns in recent year. These include the SemEval 2015 Task 11 campaign [6] on English tweets and the SENTIPOLC @ Evalita campaigns in their 2014 and 2016 editions on Italian tweets [7, 8]. Followed by, DEFT sets up a first evaluation campaign around these themes for French. The task is open to academic and industrial research teams [9]. To date, the existing methods have relied on various computational linguistics knowledge, external resources such as dictionaries and ontologies and various feature engineering mechanisms [3, 6, 8]. Overall the approaches have used heuristics in language. To entirely avoid this, we used bag-of-words (BoW's) based deep learning in DEFT 2017 shared tasks.

2 Task Description

For this challenge, three tasks of analysis of the tweets centered on the analysis of opinion and the figurative language are proposed. The challenge is divided into three tasks of increasing complexity.

Task 1: Classification of non-figurative tweets according to their polarity

Task 2: Identification of the figurative language.

Task 3: Classification of figurative and non-figurative tweets according to their polarity

The corpus for all 3 shared tasks has been given by DEFT 2017 task organizers. The detailed statistics of the corpus is displayed in Table 1

Shared task	Number of tweets in Training	Number of tweets in Testing
Task1	3906	976
Task 2	5853	1464
Task 3	5119	1281

TABLE 1 : Number of tweets taken for training and testing

3 Methodology

This section discusses methodology employed for Opinion analysis and figurative language in twitter.

3.1 Bag-of-words based system for Analysis of opinion and the figurative language on Twitter tweets

At beginning stage, we have implemented a system for analysis of opinion and the figurative language on Twitter tweets using bag-of-words (Bow's) model. Based on the domain knowledge, we set 50 as word length and embedding size to 256. Each word in tweet is turned to 256 dimensional vectors. We constructed an input matrix of dimension 3906×50 for task 1, 5853×50 for task 2, and 5119×50 for task 3. Next, we replaced each word with word embedding (categorical word representation) randomly. This forms an input tensor of shape $3906 \times 50 \times 256$ for task 1, $5853 \times 50 \times 256$ for task 2, and $5119 \times 50 \times 256$ for task 3. The maximum value in word length of 50 is chosen by using the max pooling approach. This has lessened an input tensor to matrix of size 3906×256 for task 1, 5853×256 for task 2, and 5119×256 for task 3. We typically termed a newly formed matrix as tweet embedding and we pass this matrix to logistic regression classifier.

3.2 Recurrent neural network (RNN) based system for Analysis of opinion and the figurative language on Twitter tweets

Recurrent neural network (RNN) have established as an appropriate model for sequence data modeling and that have shown remarkable performance in many of natural language processing (NLP) tasks. They are same as feed-forward networks (FFN) with an additional cyclic loop [11]. This cyclic loop carries out information from one time-step to another. As a result, RNN are able to learn the temporal patterns, value at current time-step is estimated based on the past and present states. RNN have achieved a significant performance in long standing AI tasks; machine translation, language modeling and speech recognition [10]. Generally RNN takes input as $x_t \in R^q$ and $h_{t-1} \in R^p$ of arbitrary length to compute succeeding hidden state vector h_t by using the following formulae recursively.

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

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

Where f is the nonlinear activation function, particularly logistic sigmoid function (σ) applied on element wise, h_0 is usually initialized to 0 at time-step $t0$ and $w_{xh} \in \mathbb{R}^{p \times q}$, $w_{hh} \in \mathbb{R}^{p \times p}$ and $b \in \mathbb{R}^m$ are arguments of affine transformation. Here o is the output at time step t.

We implemented RNN based system using Tensor flow for analysis of opinion and the figurative language on Twitter tweets. By following the aforementioned mechanism, we formed an input tensor of shape $3906 \times 50 \times 256$ for task 1, $5853 \times 50 \times 256$ for task 2, and $5119 \times 50 \times 256$ for task 3. Tweet embedding of each tweet of shape 50×256 is lessened to 256 dimension embedding vectors. This is passed to RNN layer to obtain optimal feature represenation. Finally RNN layer has followed by logistic regression and Arg-max function for classification.

3.3 Long short-term memory (LSTM) based system for Analysis of opinion and the figurative language on Twitter tweets

During Training RNN generates the vanishing and exploding gradient problem in the case of keeping long-term dependencies. To lighten this, [13] introduced long short-term memory (LSTM). LSTM introduced a memory block instead of a simple RNN unit. A memory block is a subnet of LSTM architecture that contains one or more memory cell with a pair of adaptive multiplicative gates as input and output gate. A memory block houses 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 is 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. The newly proposed architecture has performed well in learning long-range temporal dependencies in various artificial intelligence (AI) tasks [10]. 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 below 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 $t = 0$ hidden and memory cell state vectors such as h_0 and c_0 are initialized to 0. LSTM based system for Analysis of opinion and the figurative language on Twitter tweets is implemented using Tensor flow [12]. It has followed the aforementioned mechanism of RNN by simply replacing them with LSTM layer.

3.4 Gated recurrent unit (GRU) based system for Analysis of opinion and the figurative language on Twitter tweets

As from the above formulae, we can say that LSTM augmented with complex set of processing units. Further the research on LSTM, [14] introduced Gated recurrent unit (GRU). GRU has less number of units in compared to LSTM, computationally efficient. The mathematical formulae of GRU is given below

$$i\_f_t = \sigma(w_{xi\_f} x_t + w_{hi\_f} h_{t-1} + b_{i\_f}) \text{ (Update gate)} \tag{9}$$

$$f_t = \sigma(w_{xf} x_t + w_{hf} h_{t-1} + b_f) \text{ (Forget or reset gate)} \tag{10}$$

$$m_t = \tanh(w_{xm} x_t + w_{hm}(fr \odot h_{t-1}) + b_m) \text{ (Current memory)} \tag{11}$$

$$h_t = f \odot h_{t-1} + (1 - f) \odot m \text{ (Updated memory)} \tag{12}$$

Formulae shows, unlike LSTM memory cell with a list of gates (input, output and forget), GRU only consist of gates (update and forget) that are collectively involve in balancing the interior flow of information of the unit. In GRU, input gate (i) and forget gate (f) are combined and formed a new gating units called update gate (i_f) that mainly focus on to balance the state between the previous activation (m) and the candidate activation (h) without peephole connections and output activations. The forget gate resets the previous state (m). GRU networks looks simpler than LSTM with required only less computations. GRU based system for Analysis of opinion and the figurative language on Twitter tweets is implemented using Tensor flow [12]. It has followed the aforementioned mechanism of LSTM by simply replacing them with GRU layer.

4 Experiment and Results

We trained all experiments of various deep learning architectures using Tensorflow [12].

4.1 Cross-validation performance

We had done 10-fold cross-validation to find out optimal parameters for tweet length and embedding size. The accuracy of 10-fold cross-validation across varied tweet length and embedding size is shown in Fig 1. In this we employed Bow, RNN, LSTM and GRU network for Analysis of opinion and the figurative language on Twitter tweets. From Fig 1, we can infer that the maximum accuracy is obtained for tweet length of 50 and embedding size 256.

4.2 DEFT 2017 shared tasks results

We have submitted 3 runs for each task; run1 is based on LSTM mechanism, run2 is based on RNN mechanism and run3 is based on GRU mechanism. The detailed evaluation results has been given by the DEFT 2017 organizing committee are displayed in Table 2, Table 3 and Table 4. Table 2 has macro f-score of all participants. The employed methods performed well in Task 2 in compared to Task 1 and Task 2.

Shared task	Number of participants	Average	Median	Standard deviation	Min	Max
Task1	12	0.475	0.523	0.123	0.239	0.65
Task 2	11	0.694	0.72	0.096	0.476	0.783
Task 3	9	0.473	0.519	0.114	0.232	0.594

TABLE 2 : macro f-score for all participants

Shared task	objective			mixed			positive			negative			Macro precision	Macro recall	Macro f-score
	TP	FP	FN	TP	FP	FN	TP	FP	FN	TP	FP	FN			
T1 run1 LSTM	204	257	207	13	70	111	6	24	117	155	247	163	0.296	0.284	0.276
T1 run 2 RNN	169	214	242	2	48	122	2	25	121	182	334	136	0.227	0.254	0.228
T1 run 3 GRU	105	126	306	2	24	122	2	8	121	234	475	84	0.265	0.256	0.21
T3 run 1 LSTM	54	73	376	5	49	153	4	19	121	480	597	88	0.284	0.259	0.22
T3 run 2 RNN	71	86	359	6	65	152	4	19	121	462	568	106	0.29	0.262	0.232
T3 run 3 GRU	74	96	356	7	69	151	3	15	122	456	561	112	0.286	0.261	0.231

TABLE 3 : summary of test results respective of run in Task 1 and Task 3

Shared task	figurative			nonfigurative			Macro precision	Macro recall	Macro f-score
	TP	FP	FN	TP	FP	FN			
T2 run 1 LSTM	239	500	249	476	249	500	0.49	0.489	0.475
T2 run 2 RNN	218	477	270	499	270	477	0.481	0.479	0.470
T2 run 3 GRU	217	465	271	511	271	465	0.486	0.484	0.476

TABLE 4 : summary of Test results respective of run in Task 2

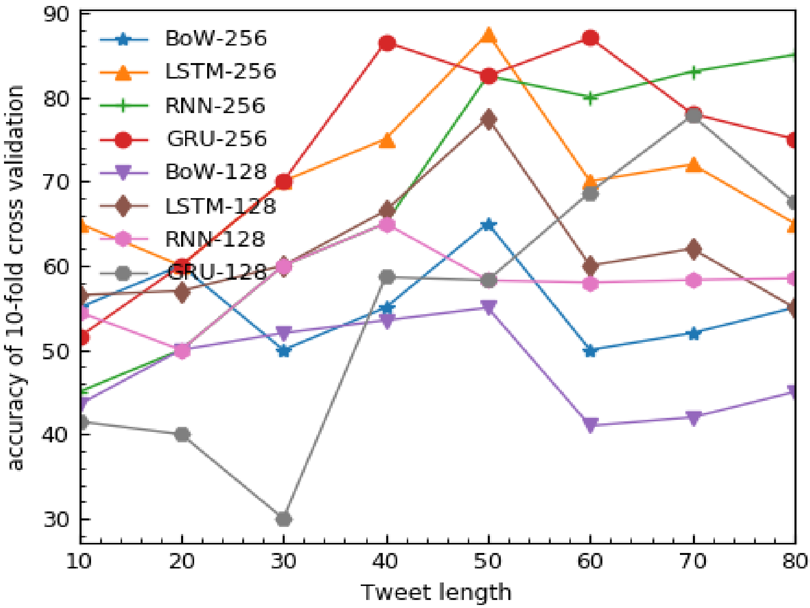

FIGURE 1 : 10-fold cross validation with Tweet length [10-80] and embedding size [128, 256]

5 Conclusion

The working note has presented a language independent method for the DEFT 2017 shared tasks such as analysis of opinion and the figurative language on Twitter tweets in French using BoWs and embedding's of RNN, LSTM and GRU. The presented supervised learning method has not relied on any resources; semantic resources such as dictionaries and ontologies or computational linguistics or feature engineering mechanisms for sentimental analysis in twitter tweets. Due to the less training corpus, the efficacy of RNN in analysis of opinion and the figurative language on Twitter tweets trails the classical BoWs approach. Though the efficacy of embedding's of RNN, LSTM and GRU is acceptable and paves the manner in future to use for the analysis of opinion and the figurative language on Twitter tweets. Evaluating the performance of RNN, LSTM and GRU embedding's with more training corpus for justification will be remained as one direction towards future work.

References

[1] PANG, B., & LEE, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, *2*(1–2), 1-135.

[2] LIU, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.

[3] MAYNARD, D., & GREENWOOD, M. A. (2014, May). Who cares about Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis. In *LREC* (pp. 4238-4243).

[4] GRICE, H. P., COLE, P., & MORGAN, J. L. (1975). Syntax and semantics. *Logic and conversation*, *3*, 41-58.

[5] CLIFT, R. (1999). Irony in conversation. *Language in society*, *28*(04), 523-553.

[6] GHOSH, A., LI, G., VEALE, T., ROSSO, P., SHUTOVA, E., BARNDEN, J., & REYES, A. (2015, June). Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)* (pp. 470-478).

[7] BASILE, V., BOLIOLIi, A., NISSIM, M., PATTI, V., & ROSSO, P. (2014). Overview of the evalita 2014 sentiment polarity classification task. In *Proceedings of the 4th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA'14)*.

[8] Barbieri, F., Basile, V., Croce, D., Nissim, M., Novielli, N., & Patti, V. (2016, December). Overview of the EVALITA 2016 SENTiment POLarity Classification Task. In *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016)*.

[9] BENAMARA, F., GROUIN, C., KAROUI, J., MORICEAU, V., ROBBA, I. Analyse d'opinion et langage figuratif dans des tweets : présentation et résultats du Défi Fouille de Textes DEFT2017. Actes de l'atelier DEFT de la conférence TALN, 26 juin 2017, Orléans, France.

[10] LIPTON, Z. C., BERKOWITZ, J., & ELKAN, C. (2015). A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*.

[11] ELMAN, J. L. (1990). Finding structure in time. *Cognitive science*, *14*(2), 179-211.

[12] ABADI, M., BARHAM, P., CHEN, J., CHEN, Z., DAVIS, A., DEAN, J., ... & KUDLUR, M. (2016, November). 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*.

[13] HOCHREITER, S., & SCHMIDHUBER, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

[14] CHO, K., VAN MERRIENBOER, B., GULCEHRE, C., BAHDANAU, D., BOUGARES, F., SCHWENK, H., & BENGIO, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

[15] PERRIN, A. (2015). Social media usage. *Pew Research Center*.