2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC 2020)

Short-term Traffic Flow Prediction Based on ConvLSTM Model

Xiaoyu Chen , Xingsheng Xie , Da Teng
Department of Automation, University of Science and Technology of China
Hefei, China
ginachen@mail.ustc.edu.cn, xshxie@ustc.edu.cn, Tengda24@mail.ustc.edu.cn

***Abstract*—This paper proposes a estimation model based on Convolutional Long Short Term Memory (ConvLSTM) model to estimate short-term traffic flow. ConvLSTM is an improved algorithm based on Long Short Term Memory (LSTM) Network. It not only establishes timing characteristics like traditional LSTM models, but also depicts local spatial features like Convolutional Neural Network (CNN). The input data is extracted from the paperual operation of CNN model at the bottom, and LSTM is used instead of the pooled layer in CNN. The space-time feature of the data is further excavated, and the complete prediction model is output through the full connection layer. Experiments show that the prediction accuracy of The ConvLSTM model is higher than that of LSTM, Two-layer LSTM and the bidirectional LSTM model.**

Keywords—ConvLSTM; CNN; LSTM; short-term traffic flow forecasting; intelligent traffic

I. Introduction

With the improvement of people's quality of life, the number of motor vehicles is increasing which brings a series of congestion problems and security risks. Effective forecasting of traffic flow can not only guide traffic police to better ease traffic congestion, but also help drivers to choose the best route in advance [1]. Traffic flow forecast can be divided into long-term and short-term traffic flow forecast. Short-term traffic prediction has become a research hotspot of intelligent traffic control. However, due to the weather, air quality, road maintenance conditions and other uncertain external factors, effective short-term traffic flow forecasting is still difficult [2].

The current short-term traffic flow forecasting model is based on three main types: time series based, statistical probability based, and machine learning based [3]. Pang used the LSTM model based on deep learning to predict traffic flow, but the method did not make full use of other parameters, only the number of learning slots, there is still a lot of room for improvement [4]. Li predicted traffic flow based on the dimensional weight residual LSTM model, which has a good generalization ability for the data in each time period, but the limitation is that the graininess obtained by the prediction is not fine enough [5]. Zhu used the BP-Adaboost model in machine learning to make short-term traffic predictions with full-day and early peaks as entry points, respectively. However, the number of iterations of the AdaBoost algorithm is not well set. It needs to be determined by means of cross-validation, and the training comparison is time-consuming and sensitive to outliers [6].

CNN is good at extracting local features, while LSTM is good at extracting sequence features [7]. In this paper, the two are combined, using CNN to extract local features as the input of LSTM for the prediction of short-term traffic flow. The results show that the model can better capture the characteristics of space-time, and its effect is better than that of the simple LSTM, Two-Layer LSTM and Bi-LSTM model.

II. MODELING PRINCIPLE

A. CNN Model

CNN is a deep neural network model consisting of convolutional layer, pooled layer, and fully connected layer, which is a weight-sharing, non-fully connected neural network, and is often used in the field of image processing [8].

Take the CNN model of 2 convolution layers and 2 pooled layers as an example as is shown in Fig. 1 [9].

C1 layer: first convolution, assuming that it enters a matrix of 32×32, through a convolution of a convolution core to generate a 28×28 size feature maps:

$$\begin{aligned} &C_i^1 = conv2(A, K_i^1, valid) + b_i^1 \\ &u_i^1 = C_i^1 \\ &a_i^1 = f(u_i^1) \end{aligned} \tag{1}$$

Where, $conv2(A, K_i^1, valid) + b_i^1$ represents a narrow convolution, f is the activation function.

S2 layer: first pooling, pooled window for 2×2, the size of 28×28 of the feature map pooled into 14×14 pool map, a total of pool maps generated:

$$\begin{aligned} &S_i^2 = \beta_i^2 down(a_i^1) + b_i^2 \\ &u_i^2 = S_i^2 \\ &a_i^2 = f(u_i^2) \end{aligned} \tag{2}$$

Where, $S = down(C)$ represents the subsampling, generally order: $S = \beta down(C) + b$, where β and b are the scalar parameters.

C3 layer: Convolution again, each 14×14 of the feature map in this layer is made up of all pool maps in the layer C_i^3 is made up of all F_1 pool maps in the pool map through a convolution core $K_{ij}^3(j=1,2,...,F_1)$, generating a total of 10×10 size sprofs of the feature maps:

$$C_i^3=\sum_{j=1}^{F_1} conv2(a_j^2,k_{ij}^3,valid)+b_{ij}^3$$
$$u_i^3=C_i^3 \tag{3}$$
$$a_i^3=f(u_i^3)$$

S4 layer: pooled again, pooled window for 2×2 , pooling a feature map of size 10×10 into a 5×5 pool map, resulting in a pool map of F_3 :

$$S_i^4=\beta_i^4 down(a_i^3)+b_i^4$$
$$u_i^4=S_i^4 \tag{4}$$
$$a_i^4=f(u_i^4)$$

Full connection layer: Finally expands the sequence $a_i^4(i=1,2,...,F_3)$ into a vector and connects in an orderly manner into a long vector as input to the full-connection layer network.

B. LSTM Model

LSTM is a commonly used recursive neural network (RNN), which can exploit the pattern of timing changes in time series such as relatively long intervals and delays. Due to the limitations of the gradient disappearance of RNN, LSTM introduced the concept of cell (cell), which is actually equivalent to the introduction of a block instead of the hidden layer inside the RNN [10]. As is shown in Fig. 2:

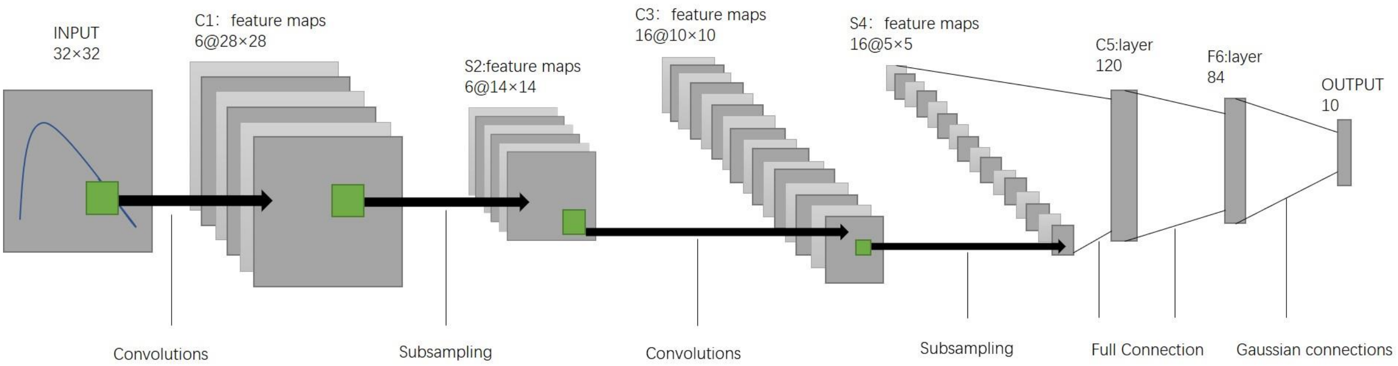

Fig. 1. The structure of CNN model of 2 convolution layers and 2 pooled layers.

Cell stores state parameters, input gate controls whether the current feature is useful, the output gate controls whether the current information is useful, and forget gate forgets the previous invalid information selection.

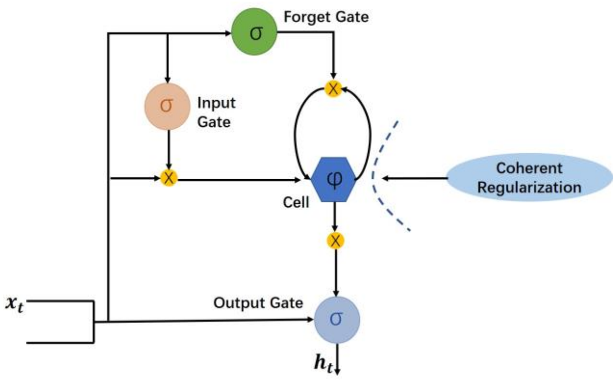

Fig. 2. The structure of LSTM model.

Input gate:

$$a_l^t=\sum_{i=1}^{I} w_{i_l}x_i^t+\sum_{h=1}^{H} w_{h_l}b_h^{t-1}+\sum_{c=1}^{C} w_{cl}s_c^{t-1} \tag{5}$$
$$b_l^t=f(a_l^t)$$

Where, a_l^t represents the sum that is accumulated by the weight of all features, b is the value that is added up and obtained by the activated function.

$\sum_{i=1}^{I} w_{ij}x_i^t$ is the input, w_{ij} represents the weight, x_i^t represents the characteristics of t ; $\sum_{h=1}^{H} w_{hi}b_h^{t-1}$ is the information hidden layer at the last moment; $\sum_{c=1}^{C} w_{ci}s_c^{t-1}$ is the information stored by the cell at the last moment.

Forget gate:

$$a_\varnothing^t=\sum_{i=1}^{I} w_{i\varnothing}x_i^t+\sum_{h=1}^{H} w_{h\varnothing}b_h^{t-1}+\sum_{c=1}^{C} w_{c\varnothing}s_c^{t-1} \tag{6}$$
$$b_\varnothing^t=f(a_\varnothing^t)$$

Output gate:

$$a_w^t = \sum_{i=1}^{I} w_{iw} x_i^t + \sum_{h=1}^{H} w_{hw} b_h^{t-1} + \sum_{c=1}^{C} w_{cw} s_c^{t-1}$$
$$b_w^t = f(a_w^t) \tag{7}$$

C. ConvLSTM Model

Considering that the extraction of data features is directly related to the upper limit of the training model, the key to the training model is the extraction of features. CNN specializes in mining the hidden local characteristics of data and the characteristics of space-time, while LSTM is typical of dealing with the problem of timing class prediction. Combining the advantages of the two, the ConvLSTM model is proposed. The effective feature skewed is extracted as a feature vector by CNN's powerful feature extraction ability, and LSTM is used instead of the pooled layer in the traditional CNN model to predict the short-term traffic flow. The network structure of the model is shown in Fig. 3:

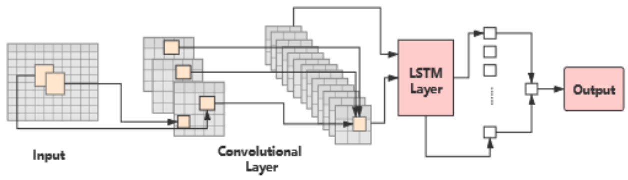

Fig. 3. The structure of ConvLSTM model.

III. CASE STUDY

A. Data Sources

Data of this experiment comes from traffic flow data obtained by detector 403324 on the I280-S Highway in Millbury, California, U.S., in 2017. The data is timed from April 1, 2017 to May 26, 2017, with a sample interval of 5 min. In order to verify the predictive effect of the model and to consistent the results of the experiment, this paper is for the traffic flow data for these eight weeks of working days, the first six weeks of weekday traffic flow data as a training set. Traffic traffic data for the last two weeks of the week is used as a test set. Use the traffic flow for the first 5 moments of a certain moment as input, and the current traffic as the output. A total of 8616 data were set and 2872 data were tested. Data features and descriptions are shown in Table 1.

TABLE I. DATA FEATURES AND DESCRIPTIONS

Feature	Description
5 Minutes	Sample interval
Flow (Veh/5 Minutes)	Traffic flow at sampling moment
Occupancy (%)	Lane occupancy
Lane Points	Lane No.
Observed (%)	Percentage of observers

B. Data Preprocessing

1) *Data monitoring learning:* The LSTM portion of the ConvLSTM model in this paper contains inputs and outputs. The traffic flow for the first 5 moments of moment T which is $T(i=1,2,3,4,5)$, can be taken as input, and the traffic flow of moment T is output as shown in Table 2.

TABLE II. INPUT AND OUTPUT OF TRAINING MODEL

INPUT	OUTPUT
$X_{T-i}(i=1,2,3,4,5)$	X_T

2) *Smoothing data difference:* Traffic flow data in the dataset is not a smooth time series, so the data differential softens. Set the original input stream data to, then

$$X^d(t) = X(t) - X(t-d) \tag{8}$$

$$\hat{X}(t-1) = \hat{X}^d(t+1) - \hat{X}(t-d+1) \tag{9}$$

Where, d is the delay time for the time series and $X^d(t)$ is the new input that predicts the value of the $\hat{X}^d(t+1)$ moment. The forecast value for the next moment $\hat{X}^d(t+1)$ can be obtained through training, and the final prediction is:

$$\hat{X}(t-1) = \hat{X}^d(t+1) - \hat{X}(t-d+1) \tag{10}$$

3) *Data generalization:* Normalizing the data so that the data range is between $(-1,1)$:

$$X_{std} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{11}$$

Where, X_{std} is the normalized data, $X_{\max}$ and $X_{\min}$ are the maximum and minimum values of the data in the time series, respectively. Eventually, the prediction needs to be denaturaled.

IV. EXPERIMENT EVALUATION

A. Training Process of ConvLSTM

The training process of the ConvLSTM model is divided into two processes. The first is CNN model training, the training of the data set as CNN input, into multiple convolution layers for the extraction of feature vectors. The result of the convolution is then entered into the LSTM layer instead of the traditional pooled layer, and the pooled results are then fully connected. The end of pooling means that the model is trained.

B. Evaluation Metrics

The evaluation metric is an important part of determining the performance of a model. This experiment uses four evaluation methods, namely: mean absolute error (MAE),root mean square error($RMSE$), mean absolute percentage error ($MAPE$), and correlation coefficient (R).

1) MAE *:* Ranging from 0 to positive infinity, which is equal to 0 when the predicted value is exactly matched with the true value. The smaller the value, the smaller the error.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{12}$$

2) RMSE : The square root error is the square root of the square and observation n ratio of the deviation between the predicted value and the real value. The smaller the value, the smaller the deviation [10].

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2} \tag{13}$$

3) MAPE : Also known as the absolute deviation of the average, is the average of the absolute value of the deviation of all individual observations from the arithmetic mean [10]. The average absolute error avoids the problem of the error canceling each other, so it can accurately reflect the size of the actual prediction error.

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{14}$$

4) R : It measures the linear correlation between the predicted value and the real value. The closer to 1,the more relevant [10].

$$R = \frac{\sum_{i=1}^{N}(y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \overline{y})^2(\hat{y}_i - \overline{\hat{y}})^2}} \tag{15}$$

In the formulae above, N is the number of samples, y_i is the actual value, $\hat{y}_i$ is the predicted value, $\overline{y}$ is the mean of the actual value, and $\overline{\hat{y}}$ is the average estimated traffic flow of N samples.

C. Experiment Result and Analysis

The prediction effect of the traffic flow in the four models in the experiment is as follows. The data from the test set is huge, and to make the comparison more obvious, we intercepted 288 data from May 15, 2017 to show the predictions as shown in Fig. 4-7:

We can see that the single LSTM model is the worst predictor, and its deviation is the largest. The Two-Layer LSTM model is slightly better predicted because it has one more layer than the LSTM model. The prediction effect of the Bi-LSTM model is close to that of the Two-Layer LSTM model. The model presented in this paper: ConvLSTM's prediction effect is significantly higher than that of the other three models, and its prediction effect is very close to the real value.

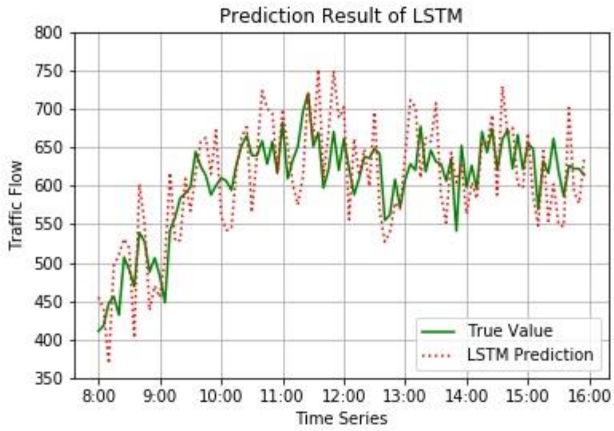

Fig. 4. The prediction result of LSTM model.

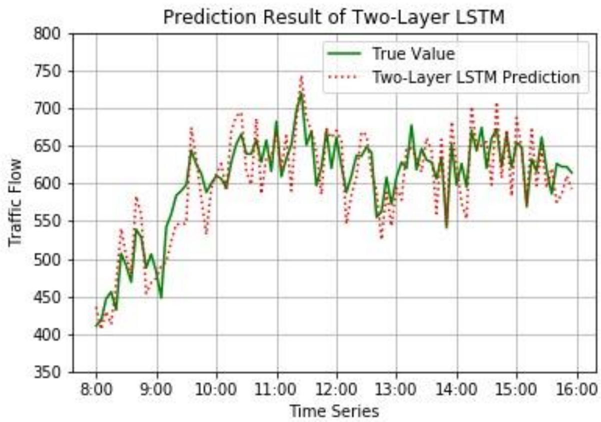

Fig. 5. The prediction result of Two-Layer LSTM model.

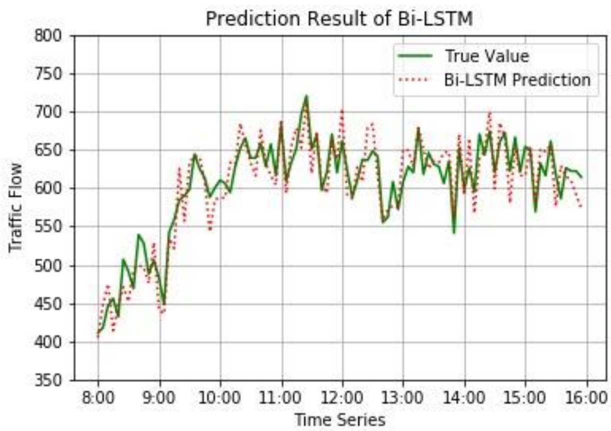

Fig. 6. The prediction result of Bi-LSTM model.

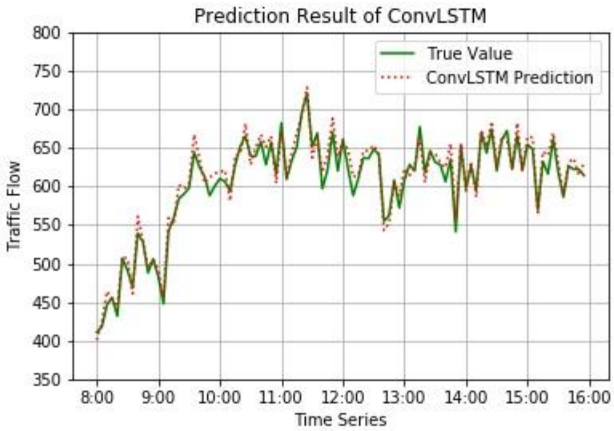

Fig. 7. The prediction result of ConvLSTM model.

To see more intuitively the difference skewed the prediction effect, we created an effect graph of the deviation of the predicted value from the real value, as shown in Fig. 8. ConvLSTM's full prediction figure for the day of May 15 is shown in Fig. 9.

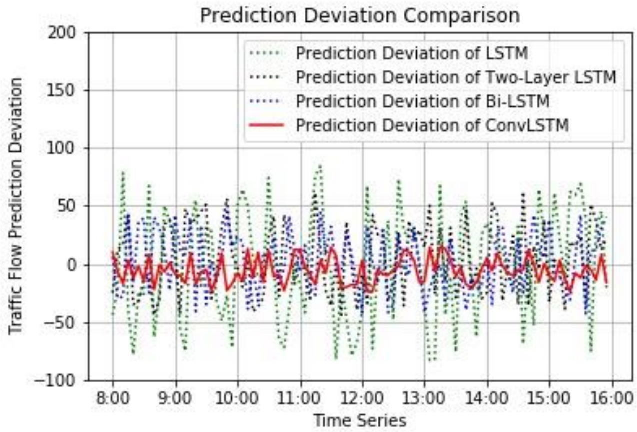

Fig. 8. Comparison of prediction deviation among four models.

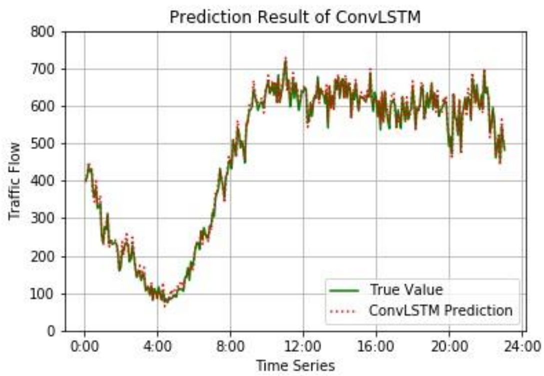

Fig. 9 .Complete prediction result of ConvLSTM model on May 15.

In addition, the relevant evaluation indicators are shown in Table 3. As you can see from the table, The prediction effect of the individual LSTM model is the worst of the four models, with the most error slots and the worst fit effect. The Two-Lsyer model is similar to the prediction effect of the Bi-LSTM model, and is slightly better than the LSTM model. The CovLSTM model has the smallest MAE and RMSE, and the most well-fitting R value, the prediction effect is the best of these four models.

TABLE III. EVALUATION RESULT OF DIFFERENT METRICS

Model	MAE	MAPE	RMSE	R
LSTM	24.2537	9.47%	21.8436	0.9379
Two-Layer LSTM	23.0829	9.61%	19.7428	0.9425
Bi-LSTM	24.0588	8.42%	18.7487	0.9401
ConvLSTM	**21.9084**	**7.13%**	**17.5439**	**0.9826**

V. CONCLUSION

Based on the short-term traffic flow forecast, this paper combines CNN with the LSTM model, and proposes a ConvLSTM model, which replaces the pooled layer in the traditional CNN model with the LSTM layer. The experiment results show that the model has higher prediction accuracy than LSTM model, Bi-LSTM model and ConvLSTM model, and is an effective method for short-term traffic flow prediction. In order to make the experimental results more consistent, without considering the influence of double-holiday and seasonal factors, the next step will be to take these two factors into account to improve the generalization ability of the model.

REFERENCES

[1] Greff, Klaus, Srivastava, Rupesh Kumar, Koutník, Jan. LSTM: A Search Space Odyssey[J]. IEEE Transactions on Neural Networks & Learning Systems, 2015, 28(10):2222-2232.

[2] Qian Chen, Xiaodan Zhu, Zhen-Hua Ling. Enhanced LSTM for Natural Language Inference[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017.

[3] Alexandre Alahi. "Social LSTM: Human Trajectory Prediction in Crowded Spaces." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE, 2016.

[4] Y. Xing, X. Ban and R. Liu, "A Short-Term Traffic Flow Prediction Method Based on Kernel Extreme Learning Machine," 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, 2018, pp. 533-536.

[5] Liu, Jun, et al. Spatio-Temporal LSTM with Trust Gates for 3D Human Action Recognition. Computer Vision – ECCV 2016. 2016.

[6] Nordin Saad, Muhammad Irfan, Rosdiazli Ibrahim.Condition Monitoring And Faults Diagnosis Of Induction Motors Electrical Signature Analysis[M].2018.

[7] L. Zhang, S.-G. Huang, Z.-X. Shi,. CAPTCHA recognition method based on RNN of LSTM[J]. 2011, 24(1):40-47.

[8] Zheng Zhao;Weihai Chena;Xingming Wu;Chen, Peter C. Y;Jingmeng Liu.LSTM network: a deep learning approach for short-term traffic forecast.[J].IET Image Processing,2017,Vol.11(1): 68-75.

[9] Gong C, Qi L, Heming L, et al. Spatio-Temporal Simulation and Analysis of Regional Ecological Security Based on Lstm[J]. 2017.

[10] Abdalraouf Hassan, and Ausif Mahmood. "Deep Learning approach for sentiment analysis of short texts." 2017 3rd International Conference on Control, Automation and Robotics (ICCAR) IEEE, 2017.