International Journal of Engineering & Technology, 7 (4.44) (2018) 47-50

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Country Information Based on Long-Term Short-Term Memory (LSTM)

Hojae Son[1]*, Anand Paul[2], Gwanggil Jeon[3]

[1]Department of Computer Science and Engineering
[2]Department of Computer Science and Engineering
[3] Department of Computer Science and Engineering
**Corresponding author E-mail: sonhj07@knu.ac.kr*

Abstract

Social platform such as Facebook, Twitter and Instagram generates tremendous data these days. Researchers make use of these data to extract meaningful information and predict future. Especially twitter is the platform people can share their thought briefly on a certain topic and it provides real-time streaming data API (Application Programming Interface) for filtering data for a purpose. Over time a country has changed its interest in other countries. People can get a benefit to see a tendency of interest as well as prediction result from twitter streaming data. Capturing twitter data flow is connected to how people think and have an interest on the topic. We believe real-time twitter data reflect this change. Long-term Short-term Memory Unit (LSTM) is the widely used deep learning unit from recurrent neural network to learn the sequence. The purpose of this work is building prediction model "Country Interest Analysis based on LSTM (CIAL)" to forecast next interval of tweet counts when it comes to referring country on the tweet post. Additionally it's necessary to cluster for analyzing multiple countries twitter data over the remote nodes. This paper presents how country attention tendency can be captured over twitter streaming data with LSTM algorithm.

Keywords: *Twitter data; Long-term short-term memory (LSTM).*

1. Introduction

Trend analysis is a common and powerful way to observe certain phenomena so that it's influential to pull out overall tendency from data. Researchers have struggle to see this pattern with a various methods such as regression, classification. Nowadays, researchers are working hard to deduce the results of these using machine learning and deep learning methods. With the availability of numerous computing powers, the validity of these techniques has proven to be certain and has had a profound impact on many areas of research. Long-term Short-term memory (LSTM) is the unit that can learn sequence property from a given data. LSTM belonged to recurrent neural network (RNN). RNN is designed to process sequential information. LSTM has feature to control forgetting, storing gates for resolving long-term dependency problem. A lot of applications can be utilized with LSTM. For example, Grammar learning, Music pattern learning, Speech Recognition, Time series prediction can be manipulated with LSTM. The common characteristic of examples above is that the model learns the property from the sequence. When human recognize a pattern from a phenomenon we discover and memorize patterns of it. From this idea RNN researchers are enthusiastic about imitating how human-being learn a pattern. Time series data prediction is one of the popular research area corresponding to that perspective. People wants to see near future based on previous data such as stock market or weather forecast. It's crucial part to get a prediction result of next time interval for real-time data with past history. Twitter is one of social platform which people can leave their thought on specific topic or their status instantly over the countries. It has generated tremendous data every day. With the perspective of data analyst it can be regarded as flow of data stream. In other words, data stream can reflect current flow on a specific topic. Twitter provides real-time streaming data generated over the world as well as filtration function based on keyword, language, geo-location and follow. Data analyst manipulates this feature for their own purpose. For instance a businessman needs to know what people need and how people think of a product. It is efficient to acquire information with stream data. This paper presents how LSTM model captures meaningful information from twitter stream data applied for machine learning technique. Especially we focus on obtaining tendency of interest between countries based on tweet data. It takes several steps in perspective of both processing data and analyzing data. Data process is indispensable to handle real-time data as soon as the system receive data which means the system should finish their work with current data before next coming data is loaded. From there, there is a need for data distribution processing with cluster. Google provides analytics how people search a particular topic based on geo-location, time through Google Trends. It is good material to predict near future as well On the other hand, twitter has a different property. It captures current trending message containing all kinds of information encompassing emotional status, happening, positive or negative opinion on a topic etc. Users over the world generates enormous twitter data endlessly. With these twitter streaming data it is possible to grasp the tendency how much a country pays an attention on the other countries. The challenge of this work does not include sentiment analysis on each sentence. It can shows the interest rate in a sense either positive or negative way.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2. Related Work

2.1. Twitter

A lot of researchers make use of social data for different purpose. In [1], authors introduce twitter is micro-blogging service on the social service platform. They describe twitter is generating knowledge discovery data continuously. With twitter data stream, they applied for learning algorithm to analyse and extract information "what's-happening-right-now" with statistic perspective. There are a couple research with "what's-happening-right-now" data. For instance, Takeshi Sakaki tries to detect earthquake in a real-time with social sensor in [2]. He suggests to build reporting system with location filtering techniques. In addition to this he uses twitter semantic analysis and sensor's value analysis to send event detection promptly. Similarly Paul S. Earle explains how useful twitter data can capture earthquake in social world in [3]. He filters the sentence containing 'earthquake' from the tweet and tries to identify earthquake with a short-term-average, long-term-average algorithm. Furthermore, they categorized earthquake into "alert dissemination", "Situational awareness" and "Event detection" to clarify what's happening there. There are another application to detect current happening with twitter data. [4] suggests 'Twitter Monitor' system to find out emerging topics for trend recognition. Firstly they detect the 'bursty' keyword which comes at a high rate for a time interval and then they make a group of those words corresponding to topics. Likewise, twitter data is useful to detect the current phenomenon. Additionally some of authors try to predict near future with the past and current generating data. Delenn Chin analyse twitter sentiment to predict who's going to be elected from 2016 Presidential Candidates tweet. He classifies tweets based on emoji and applies Support Vector machine algorithm with degree of emoji and geolocation in [5]. In similar fashion, Erik Tjong Kim Sang predicts the 2011 Dotch Senate Election. He does prediction with entity counts by putting influential parameter for the election in [6]. Harshavardhan Achrekar et al foresee flu trends with tweet. SNEFT (Social Network Enabled Flu Trends) architecture is suggested to track and predict spread of an influenza epidemic. Basically this system detects flu keyword ('Flu', 'Swine Flu','H1N1') and give a warning signal to the users belonged to the region under influence in [7]. They correlated influenza-like illness (ILI) data with information provided by the Center for disease control and prevention and shows twitter data help improving real-time evaluation of ILI activity. In [8], John bollen interprets tweet as economical behaviour which indicates decision making of human. He correlated time series twitter data and Dow Jones Industrial Average (DIJIA) prediction model. These papers say twitter social data can present symptom of any topics and predict following result based on it. In a sense of data mining over social network, Anand Paul analyze human behavior with social big data system in [9].

2.2. Forecasting with LSTM

Machine learning is one of trending technologies with various application. It is commonly productive to do classification, detection and regression. The advantage of regression algorithm with time series data is useful to grasp a tendency from data. In [10] U. Thissen introduces to predict time series data with support vector machine. [11] formulates nowcasting model with spatiotemporal sequence forecasting with convolutional LSTM. It has a modified version of LSTM algorithm but the base of the proposal begins with fully-connected LSTM. In [12] Mohammad Assaad et al also modifies recurrent neural network by adapting boosting algorithm and compare with existing methods including vanilla RNN algorithm. If there is a sequence property on data, people try to apply for LSTM to acquire next value/sequence. In [13] Xiaolei Ma et al captures travel speed with traffic microwave detectors in Beijing. Similarly [14] Zheng Zhao uses LSTM to forecast short-term traffic information and he mentions it could be beneficial to control travel modes, travel routes and so on.

3. Proposed Work

Country Interest Analysis based on LSTM (CIAL) estimates the number of destination country name referred from tweet filtered by source language. People can get a benefit when they need to understand the tendency how interest varies between countries. In other words, CIAL model detects what country gets an attention from people who use source language. It should be noted that countries that are highly correlated, such as neighboring countries, are forced to count more. So we have to focus on the rate of change, not count quantity. Fig 1 shows the overall architecture of CIAL model.

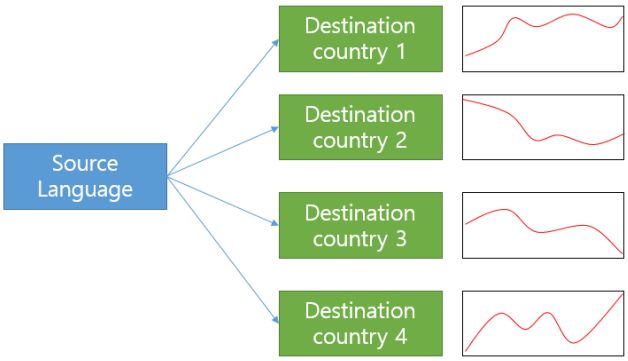


Fig. 1: Country Interest Analysis based on LSTM

It is processed in three steps. Firstly it needs to collect real-time twitter data and filters data based on source language. If tweet contains a particular country name, it adds to the counting pool. Secondly, it needs to distribute data corresponding to target country since there might be some crash unless dealing with the real-time data at a right time. Lastly it predicts the next interval of count number of countries according to the specified time period with LSTM model.

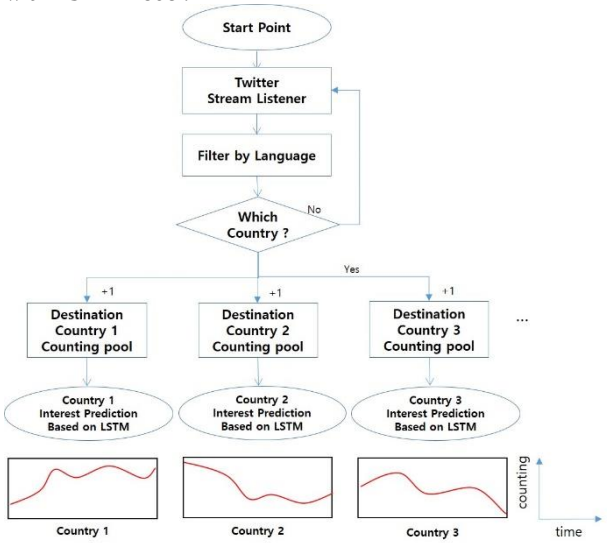


Fig. 2: Flow chart of proposed work (CIAL)

3.1. Data Processing

Twitter provides user interface to filter all of real-time streaming tweet data. They also give Http API for developers however, there is request rate limitation with http so streaming feature is much better option to deal with real-time data. As Fig 2 shows, the model filters source language and finds country name from sentence.

However there is a hurdle to process finding countries because of searching a lot of countries for each sentence. In this case twitter streaming api give protocol error with the 'Incomplete Read' message. That is why the model needs to distribute traffic to cluster.

3.2. Neural Network

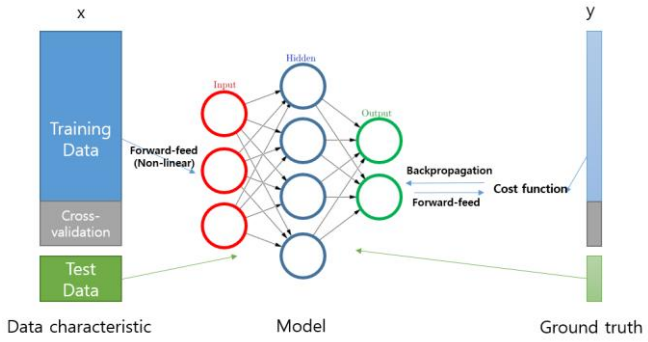


Fig. 1: Neural net overview

Neural network is specialized to optimize generic problem based on given data. There have been a tons of efforts to reflect characteristic of data on the model. With a perspective of supervised learning there are several parts that are carefully considered. For example, it's important to choose activation function since most recent problems have nonlinear property. Additionally when it comes to forward-feeding & backward-feeding, the loss of weight through network should be monitored in order not to lose training states. Depending on the problem, considering cost function is one of the issues between classification and regression. Optimization method is necessary to tune weights in a right direction. These whole issues have been investigated actively on neural network research.

3.3. Recurrent Neural network (RNN)

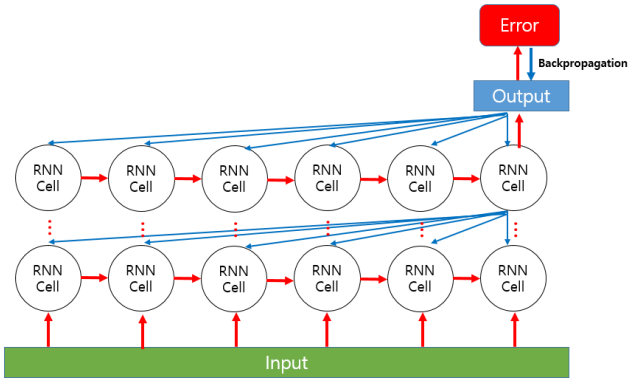


Fig. 2: Recurrent neural network (RNN)

RNN works especially with sequential data. Since this model is able to convey information from previous sequence data, RNN can interpret context information by memorizing occurrence of pattern. The problem of RNN is vanishing & exploding gradient gradually over long-term and deep neural net. If this problem occurs, there is no way to maintain learning parameter over training time. So [15] introduced Long-term short-term which is the modified version of classic RNN with ability to adjust states to determine what to store and what to forget. LSTM replaced the RNN cell with additional states to maintain gradient value.

3.4. Long-term Short-term Memory (LSTM)

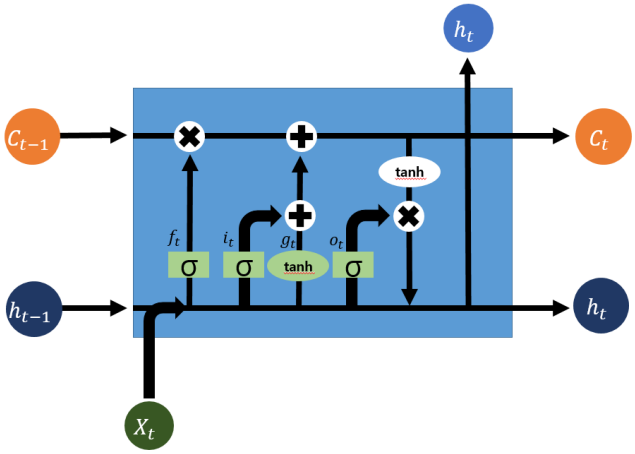


Fig. 3: Long-term Short-term Memory Architecture

The background of LSTM begins with the problem of vanishing gradient problem and exploding gradient problem on vanilla RNN. [16] Felix A. Gers et al mention it's crucial to reset internal states not to break down network. While training data, error backpropagates to adjust weights with maintaining training properties. Fig.4. is typical flow how model extracts property from sequence data. tanh function is an activation function to reflect non-linearity of data. It consists of gates, cell states, input and output. The model retain properties to remember and detain properties to forget. In f_t Previous output and current input determines what to remember based on previous memory. With i_t it control what to store from input X_t and it associate with g_t to update cell state to transfer memory over next units. Based on this memory cell and o_t it will produce next output h_t. It can be summarized following equations.

$$f_t = \sigma(W_f * x_t + U_f{*}h_{t-1} + b_f)$$
$$i_t = \sigma(W_i * x_t + U_i{*}h_{t-1} + b_i)$$
$$o_t = \sigma(W_o * x_t + U_o{*}h_{t-1} + b_o)$$
$$g_t = \tanh(W_g * x_t + U_g{*}h_{t-1} + b_g)$$
$$c_t = i_t \otimes g_t \oplus f_t \otimes c_{t-1}$$
$$h_t = o_t \otimes \tanh(c_t)$$

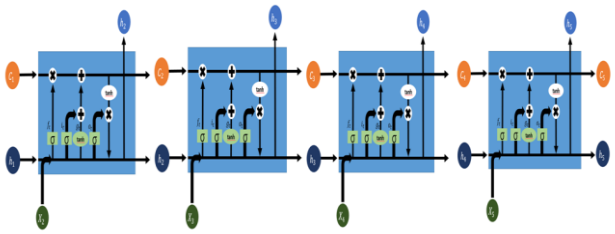


Fig. 4: LSTM Overview

The interesting thing to use LSTM model with sequential data is it can learn with various type of target. For instance language model targets to generate sentence which consists of several words/outputs. It is called many to many training model. In the case of our proposal, with sequence of previous history, it targets specific point of timeline called many to one training model. It is possible to aim different output since error propagates depending on cost function from ground truth.

4. Evaluation

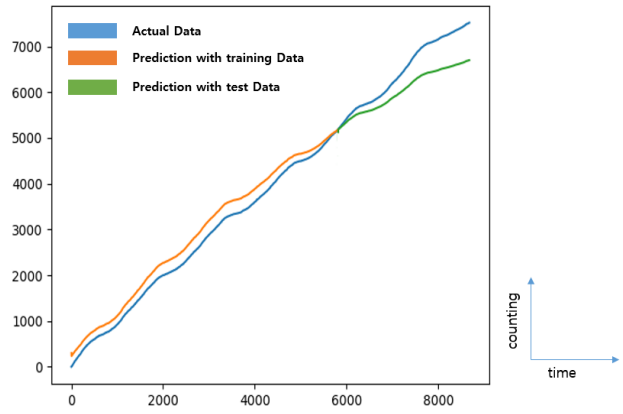


Fig. 5: Country Interest Analysis model with LSTM

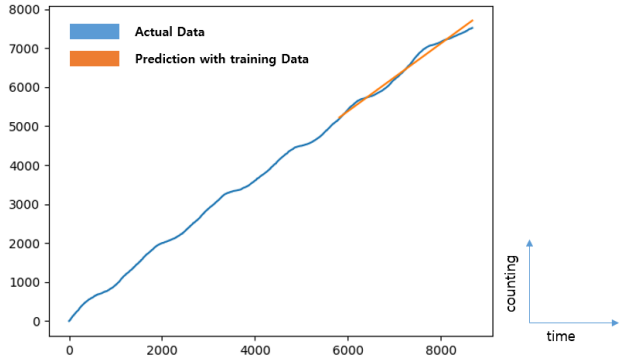


Fig. 6: Linear Regression

Table 1: Comparison between LSTM and Linear Regression

	Long-term Short-term Memory	Linear Regression
Correlation Coefficient	0.999411	0.991697

The twitter data are collected every minutes and targets for India as it generates most of data. LSTM module is executed with Mean square error and Adam optimizer. It is tested with 20 epoch. The length of training size sets to every 3 minutes. LSTM architecture units 4 units and it results in 1.9998e-05 loss. As described Fig 7 above, the test data follows shape actual data. In Fig 8 it shows linear regression with given data. It looks linear regression better compared to LSTM model however, Table 1 shows that regression rate result is slightly better which means LSTM capture tendency. Lastly the test is proceeded with tensorflow1.3.0 & python 3.5.2 with Windows 10 environment.

5. Conclusion and Future Work

Proposed scheme tries to extract sequential feature of social data and applies to Long-term Short-term memory model. We introduced overall recent neural network concept and recurrent neural network. In addition to those concepts, we presents LSTM architecture as modified version of RNN. Social data are extremely generated every day. If there are sequential property on social data, it's going to be useful to understand and capture stream of tendency. Our proposed framework uses twitter data to predict how data flow moves by counting keyword on tweet. LSTM model grasps twitter tendency with neural-net optimization approach. Our future work will be applying more data to capture more accurate evaluation. Additionally next topic will be building architecture to handle overloaded real-time social data over clusters and update neural model to reflect on-line algorithm.

Acknowledgement

This study was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-2017R1C1B5017464). This study was supported by the Brain Korea 21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (21A20131600005).

References

[1] Albert Bifet and Eibe Frank, "Sentiment Knowledge Discovery in Twitter Streaming Data" Discovery Science, 2010
[2] Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo, "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors" WWW '10 Proceedings of the 19th international conference on World wide web
[3] Paul S. Earle, Daniel C. Bowden, Michelle Guy, " Twitter earthquake detection: earthquake monitoring in a social world", Annals of Geophysics U.S. Geological Survey, Denver, CO, USA 2011
[4] Michael Mathioudakis, Nick Koudas, TwitterMonitor: Trend Detection over the Twitter Stream" Conference: Conference: Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2010, Indianapolis, Indiana, USA, June 6-10, 2010
[5] Delenn Chin, Anna Zappone, Jessica Zhao, " Analyzing Twitter Sentiment of the 2016 Presidential Candidates", American Journal Of Science and Research
[6] Erik Tjong Kim Sang and Johan Bos, "Predicting the 2011 Dutch Senate Election Results with Twitter", Proceeding Proceedings of the Workshop on Semantic Analysis in Social Media, 2012.
[7] Harshavardhan Achrekar ; Avinash Gandhe ; Ross Lazarus ; Ssu-Hsin Yu ; Benyuan Liu, " Predicting Flu Trends using Twitter Data", 2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) 2011
[8] Johan Bollen and Huina Mao , "Twitter Mood as a Stock Market Predictor", Journal of Computational Science, Indiana University Bloomington, 2011
[9] A Paul, A Ahmad, MM Rathore, S Jabbar, "Smartbuddy: defining human behaviors using big data analytics in social internet of things", IEEE Wireless communications,2016
[10] UThissen Rvan Brakel A.Pde Weijer W.JMelssen L.M.CBuydens, "Using support vector machines for time series prediction",Chemometrics and Intelligent Laboratory Systems, 2003
[11] Xingjian Shi Zhourong Chen Hao Wang Dit-Yan Yeung, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting", Computer Vision and Pattern Recognition, 2015
[12] Mohammad Assaad Romuald Boné Hubert Cardot A new boosting algorithm for improved time-series forecasting Information Fusion, 2015
[13] Xiaolei Ma, Zhimin Tao, Yinhai Wang, Haiyang Yu, Yunpeng Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data", Transportation Research Part C Emerging Technologies, 2015
[14] Zheng Zhao ; Weihai Chen ; Xingming Wu ; Peter C. Y. Chen ; Jingmeng Liu, "LSTM network: a deep learning approach for short-term traffic forecast", IET Intelligent Transport Systems, 2017
[15] Sepp Hochreiter, Jürgen Schmidhuber, "Long Short-term Memory",Neural Computation, 1997
[16] Felix A. Gers , Jürgen Schmidhuber and Fred Cummins, "Learning to Forget: Continual Prediction with LSTM",1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)