Original Article

Recognition of Traffic Information with the Help of Social Media Tweets

Dr.C. Sivakumar[1]

[1]Assistant Professor, Department of Computer Science and Systems Engineering, Sree Vidyanikethan Engineering College (Autonomous), Andhra Pradesh, India. E-mail: svkumar650@gmail.com

Abstract

Social media has seen a significant change in the current day in terms of how people use it to interact with friends and family, publish content, and share it. A few social media platforms, like Twitter, LinkedIn, and Facebook, were also utilized to share the current everyday trends in the actual world. People often struggle to get to their destinations on time in major cities with heavy traffic, like Washington DC, Beijing, and New Delhi. To ensure that travelers reach their destination or locations quickly, we concentrated on collecting useful traffic-related data from social media content in our study. Additionally, we were committed to providing immediate safety precautions for traffic incidents. The key benefit is the improvement of traffic event accuracy and the presentation of location data for heavy traffic. In this study, the CNN-LSTM (Convolution Neural Network-Long Short Term Memory) approach and the Logistic Regression method were tested for their ability to extract real-time traffic information.

Keywords: Logistic Regression, CNN, LSTM, CNN-LSTM, Social Media, Traffic Information.

DOI: 10.47750/pnr.2022.13.S03.005

Introduction

The usage of social media for posting and sharing information has increased significantly during the past ten years. Twitter is a highly well-liked microblog that has recently drawn greater notice. Many people use this social network to communicate digitally with family and friends across the world (we have approximately 145 million active users every day). Twitter is a vast platform where people share everything, with 500 million tweets being sent there every day.

Because of its popularity and data volume, we can utilize these available useful dates to develop a system that can effective analysis this raw data and consolidated the required information for the further process of final execution. In the modern-day, this extracted public information has attracted a lot of interest in a variety of areas, including the detection of natural disasters like earthquakes and fire mishaps, as well as the use of real-time information extracted from tweets.

Hence the floating social media data on the internet nowadays started playing a vital role in regulating the congested road transport system of an urban transport system that is straggling to regulate the traffic during pick hours. The above-stated methodology will be very effective adopted in the future in most dance-populated urban cities like Mumbai, Delhi, Calcutta, and Chennai where we can find more social media users in common and the will establishment of mobile network coverage makes this practical possible in real-time senior

The following summarises the paper's impact:

i. Traffic data is gleaned from social media using deep learning algorithms.
ii. Traffic event detection is based on extracted social media data to assess traffic event data about the location of the traffic incident.

Related Works

First, the researchers gathered over 1330 tweets, of which half are related to traffic and the other half are not. SVM (Support Vector Machine), MLP (Multilayer Perceptron), and Naive Bayes are included in the baseline model to categorize the texts. The SVM model performed the best during the experiments. Italian highways have the SVM system installed and being tested. When compared to local newspapers and websites, it virtually always identifies traffic in real-time.

Later, they obtained data from the Sina-Webino website and extracted the traffic data using CNN-LSTM. These models are also employed for traffic prediction.

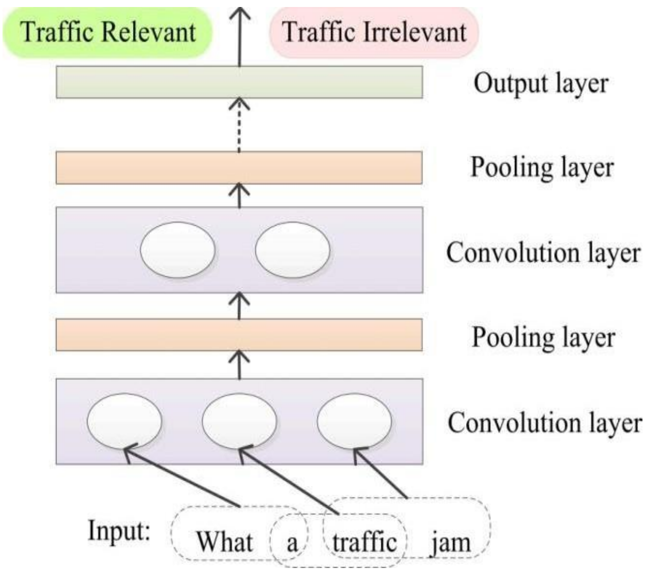

Fig. 1: CNN Architecture for Text Classification

A. CNN for Text Classification

This group of artificial neural networks is deep and feed-forward. It makes use of a multilayered perceptron variant that needs little preprocessing. The CNN architecture for text categorization is shown in Fig. 1.

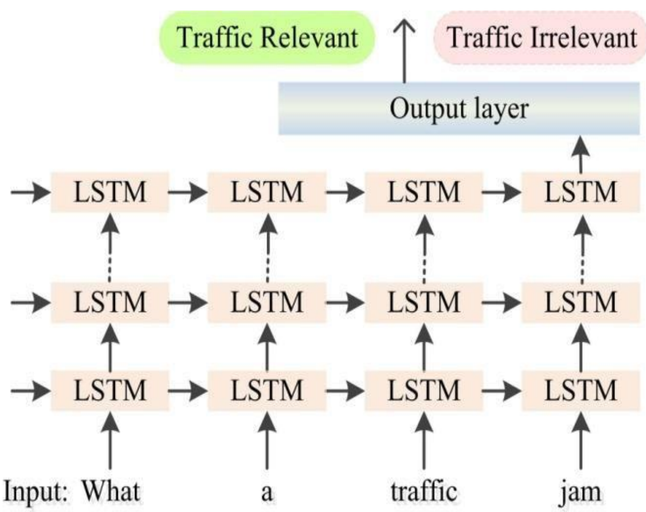

Fig. 2: LSTM Architecture for Text Classification

B. LSTM for Text Classification

The artificial recurrent neural network architecture known as LSTM (Long Short Term Memory) is mostly employed in deep learning. It is frequently used in sequence learning and primarily learns from previous experience to forecast time series.

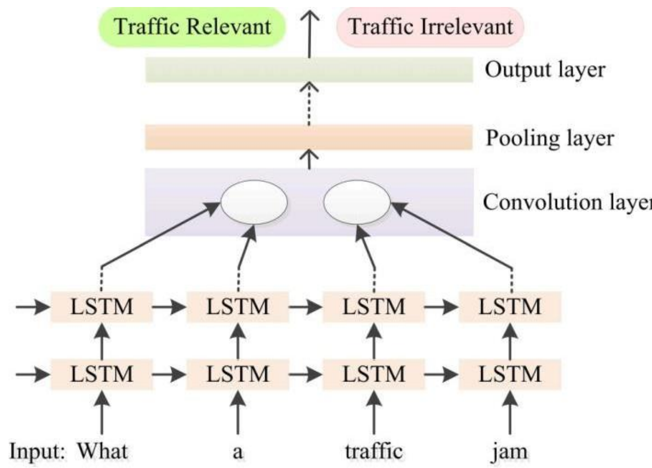

Fig. 3: CNN-LSTM Text Classification Architecture

CNN-LSTM Text Classification Architecture

The input is initially fed into the primary dictation layer, as shown in Fig. 3, and then several more LSTM layers are applied. The following layer is a convolution layer followed by a subsequent layer, and there may be more additional filtering layer pairings.

Methodology

For the text categorization in the suggested system, we use a logistic regression model. From the social media microblog known as Twitter, we have almost gathered 40,879 tweets. It has nearly three rows (class, tweet id, and tweet text). The data is trained and tested, with 60% of the data being trained and 40% of the data being manually tested. To examine traffic data, we use a binary logistic regression model. Where tweets about traffic are given the value "1," whereas tweets about unrelated traffic are given the value "0." It allows us to categorize tweets into two groups: those that are connected to traffic and those that are not.

The tweets about traffic are filtered and saved in the word cloud. Based on the tweets we have gathered from Twitter, we will identify the regions on the map where the traffic is significantly concentrated. Traffic information is categorized according to location.

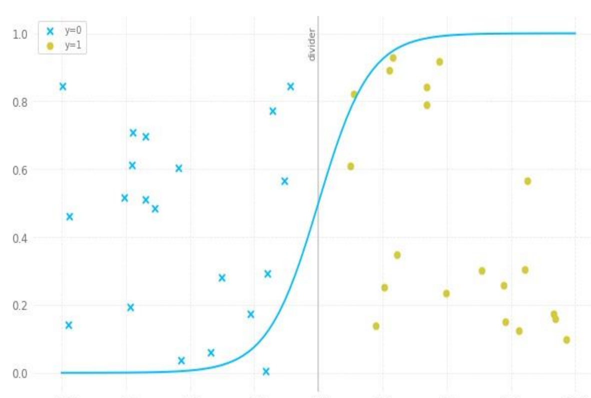

Fig. 4: Binary Logistic Regression

We categorize the social media text using binary logistic regression, as seen in the accompanying image. This division of relevant and irrelevant traffic data into two groups is done in binary format. We are computing the accuracy final result to evaluate the performance.

Table I: Performance Indexes

Index	Definition
Precision	$pre = \frac{TP}{TP+FP}$
Recall	$rec = \frac{TP}{TP+FN}$
F measure	$F_{\beta} = (1+\beta^2)\frac{pre \cdot rec}{\beta^2 \cdot pre + rec}$

We are computing the accuracy, recall, and f measure values using the references from table 1.

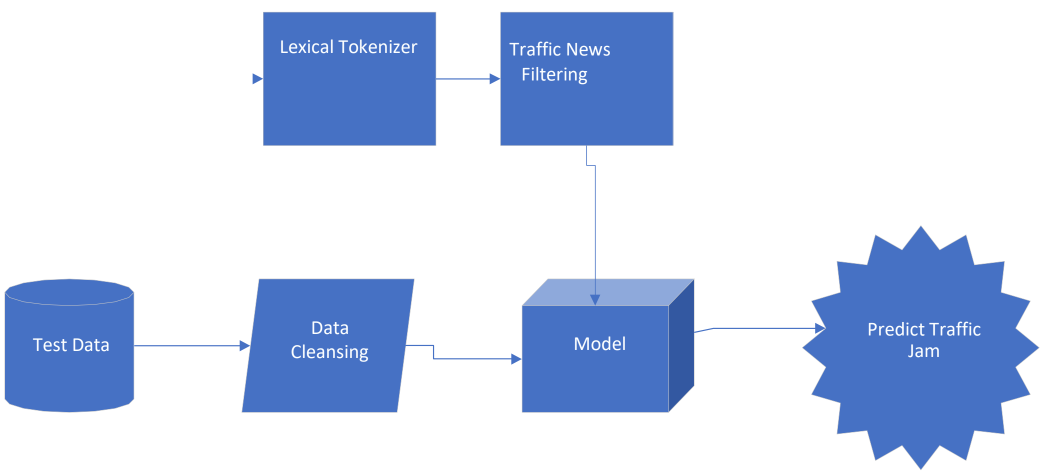

Fig. 5: System Architecture

The data used in this study was gathered from social media, and after manually filtering and training the data sets, we used the trained dataset in conjunction with a lexical tokenizer to identify stop words. By using data, we can forecast social traffic jams.

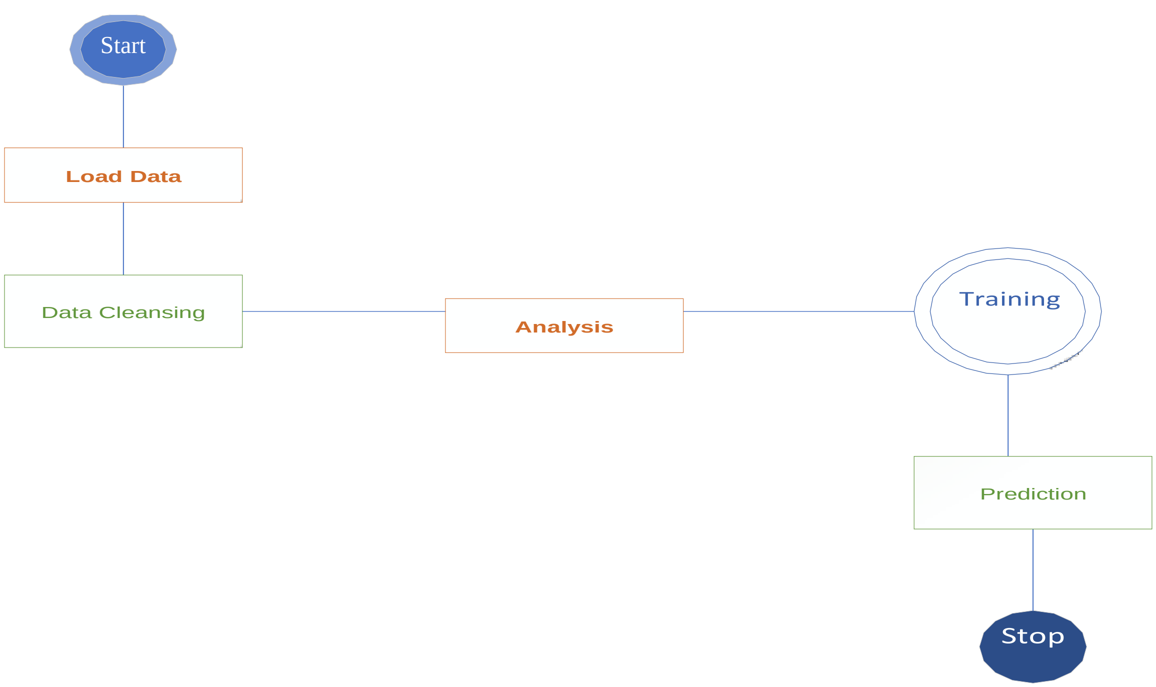

Fig. 6: Flow Chart of the System

RESULTS

Table II: Performance Comparison of LSTM-CNN and Logistic Regression

Task	Pre	Rec	F1
CNN-LSTM	0.98470	0.47234	**0.63844**
Logistic Regression	0.75528	0.75187	**0.75357**

According to the aforementioned table, Logistic Regression outperforms the CNN-LSTM model combo. With the use of table 1, we estimated the pre, rec, and f measure values.

CONCLUSION

Other deep learning techniques outperform the model we've suggested. The best proof of concept demonstrates how easy and practical traffic management can be applied to social media results. Within the sphere of traffic social media, the suggested solution identified crucial regions. The increased usage of social media for traffic management opens up several opportunities for creating solutions to serve the public who are interested in traffic news. The incorporation of data from wearable and other forms of technology will benefit these outcomes. In the future, you may gather additional information from social media platforms and use machine learning or deep learning algorithms to attain the above-discussed goals in real-time.

REFERENCES

Detecting Traffic Information From Social Media Texts With Deep Learning Approaches Yuanyuan Chen, Yisheng Lv, Senior Member, IEEE, Xiao Wang, Member, IEEE, Lingxi Li, Senior Member, IEEE, and Fei-Yue Wang, Fellow, IEEE.

Dr.S. Vasundra, CSE, JNTUACEA, Published the paper "Impact of Healthcare Predictions with Big Data Analytics and Cognitive Computing Techniques", IJRTE-International Journal of Recent Technology and Engineering. ISSN NO: 2277-3878, Volume-8 Issue-2. July 2019.

Societies 5.0: A New Paradigm for Computational Social Systems Research Fei- Yue Wang, Fellow, IEEE, Yong Yuan, Senior Member, IEEE, Xiao Wang, and Rui Qin.

Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo.

Twitter Catches The Flu: Detecting Influenza Epidemics using Twitter Eiji Aramaki Sachiko Maekawa Mizuki MORITA the University of Tokyo National Institute of JST PRESTO Biomedical Innovation Tokyo, Japan Tokyo, Japan Osaka, Japan. eiji.aramaki@gmail.com, Sachiko.maskawa@gm ail.com morita.mizuki@gmail.com.

Yue, M., Fan, L., Shahabi, C. 'Traffic accident detection with spatiotemporal impact measurement'. Proc. of Pacific-Asia Conf. on Knowledge Discovery and Data Mining, Melbourne, Australia, 2018, pp. 471–482.

Jiang Xiaorui, Tian Ya, Jiang Li, Liang Ronghua, "Visual analytics of urban road transportation data: a survey", Journal of Image and Graphics, Vol.20, No.4, 2015, 0454-0467.

F.Y. Wang, Y. Yuan, X. Wang, and R. Qin, "Societies 5.0: A new paradigm for computational social systems research," IEEE Trans. Computat. Social Syst., vol. 5, no. 1, pp. 2–8, Mar. 2018.

A.F. Abidin, M. Kolberg, & A. Hussain, Improved traffic prediction accuracy in public transport using trusted information in social networks, In Proc. 7th York Doctoral Symp. Comput. Sci. Electron., 2014, p. 19.

Ibrahim, S. (2022). Mathematical Modelling and Computational Analysis of Covid-19 Epidemic in Erbil Kurdistan Using Modified Lagrange Interpolating Polynomial. International Journal of Foundations of Computer Science, 1-17.