Predicting Incidents of Crime through LSTM Neural Networks in Smart City Domain

Ulises M. Ramirez-Alcocer
Faculty of Engineering and Science
Autonomous University of Tamaulipas
Victoria, Tamaulipas, Mexico
email: a2093010066@alumnos.uat.edu.mx

Edgar Tello-Leal
Faculty of Engineering and Science
Autonomous University of Tamaulipas
Victoria, Tamaulipas, Mexico
email: etello@uat.edu.mx

Jonathan A. Mata-Torres
Reynosa-Rodhe Multidisciplinary Unit
Autonomous University of Tamaulipas
Reynosa, Tamaulipas, Mexico
email: a2093010058@alumnos.uat.edu.mx

***Abstract*—Crimes are common social problems that affect the quality of life, economic growth and reputation of a country. In smart cities, the aim is to reduce crime rates using Information and Communication Technologies (ICT), specifically with the use of Internet of Things (IoT) technology in combination with legacy information systems, in order to obtain data automatically. In this paper, we propose an approach based on deep learning for the classification of incidents of a crime of public safety through predictive analysis. The predictive model is based on a neural network Long Short-Term Memory (LSTM), trained with a small group of attributes, enabling the prediction of the class label in the validation stage, with a high percentage of prediction accuracy. The proposed approach is evaluated through a big data set (real data) of type open data, which contains historical information about the crimes of a smart city.**

Keywords-LSTM; Prediction; Neural Network; Data Classification; Smart City.

I. INTRODUCTION

Nowadays, smart city approaches are developed in order to improve the economy, mobility, environment, people life, living standards, and governance of cities [1], supported by the design, implementation, management, and monitoring of projects based on Information and Communication Technologies (ICT). The main smart city characteristics include sustainability, resilience, governance, enhanced quality of life and intelligent management of natural resources and city facilities [2]. In this way, for the development of the services deployed in the complex systems of intelligent cities, the Internet of Things (IoT) technologies have emerged as a new computational paradigm.

The IoT consists of a large number of objects (physical and virtual things) with pervasive sensing, detection, actuation, and computational capabilities, allowing these devices to generate, exchange, and consume data with minimal human intervention [3]. Contributions from the implementation of IoT technologies can be found in smart cities, smart homes, e-health, smart grids, intelligent transportation systems, crime prevention systems, and intelligent use of water, among others. The extensive use of information systems (smart type) in various domains of the city, as well as the use of smart devices in people's daily life, have as a consequence the production of massive volumes of data [4], known as big data. These data are generated by human-to-human, human-to-machine, and machine-to-machine interactions. The big data volumes consist of a mixture of complex data characterized by large and fast-growing data sets [2], which exceed the analysis capabilities of the current data management systems. Therefore, there are requirements related to the acquisition and storage of data, the processing of information, and the development of algorithms for intelligent decision support systems. In addition, the monitoring, understanding, and analysis of the data generated by the information systems using algorithms based on data mining and machine learning, can be used to advance their contribution to the goals of sustainable development in a smart city [5]. The public safety data set of smart cities holds a large amount of crime data that could be exploited to discover patterns and predict future crime trends [6]. The predictive analysis aims to optimize the exploitation of these data in order to use knowledge discovery to anticipate criminal events [6] [7].

Deep learning techniques have demonstrated their capability of the discriminative power compared with other machine learning methods. Recurrent Neural Network (RNN) architecture has become a model of the deep learning techniques most successfully implemented in different domains, due to its natural ability to process sequential entries and to know their long-term dependencies [8]. That is, RNN is designed to utilize sequential information of input data with cyclic connections among building blocks. In the RNN, neurons are connected to each other in the same hidden layer and a training function is applied to the hidden states repeatedly [8]. The Long Short-Term Memory (LSTM) neural network is an extension of the RNN, which has achieved excellent performance in various tasks, especially for sequential problems [9]-[11]. The implementation of LSTM neural networks for the prediction of the class label (classification) from a set of instances through predictive analysis can be considered an important strategy as a technique in the context of supervised machine learning and data mining.

 ISBN: 978-1-61208-730-6

In this paper, we propose a deep learning-based predictive analysis approach for the classification of crime incidents. The predictive model is based on an LSTM neural network that is trained with a small number of attributes of the big data set, enabling the prediction of the class label in the validation phase. In order to validate the approach and show the applicability to the public safety domain, we present preliminary results using a data set with 6.9 million registers. The test carried out on the trained LSTM network shows that it has the capacity to predict the class label of a new instance. For the evaluation of the system model, we present a case study with data collected by a real smart city system in Chicago, USA. This system combines IoT devices and legacy system in smart city domain. Our case study addresses the challenges of data analytics of smart city public safety. In addition, a methodology is proposed to guide the preprocessing and categorization of the input data to the neural network, as well as the design and training of the network.

The rest of the paper is structured as follows. Section II presents a background review on LSTM neural network. Section III presents a methodology proposed to construct the predictive model. Section IV discusses the findings, while Section V we discusses the related work. Section VI provides the conclusions and plans for future work.

II. Research Proposal

An LSTM neural network is considered a network with a special structure consisting of memory blocks and memory cells, together with the gate units (input, forget, and output) that contain them [12]. This structure allows the LSTM network to select which information is forgotten or remembered (see Figure 1). The multiplicative input gate units are used to avoid the negative effects that unrelated inputs can create.

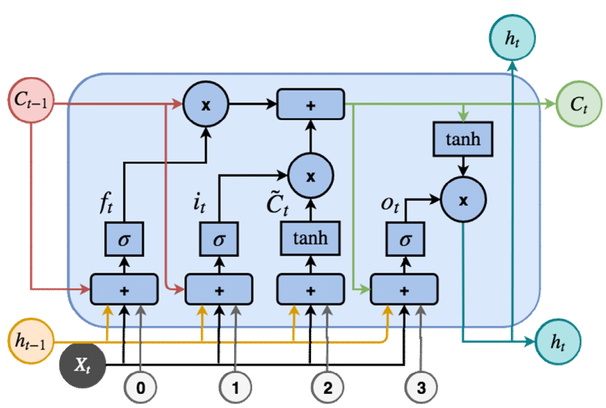

Figure 1. Structure of an LSTM cell.

In our proposed approach, the following is considered for the definition of the neural network model: 1) the *input gate* controls the input flow to the memory cell, 2) the *output gate* controls the output sequence of the memory cell to other LSTM blocks, and 3) the *forget gate* in the structure of the memory block is controlled by a single-layer of the neural network, as shown in Figure 1.

The operation of the LSTM network is based on equations (1), (2), (3), (4), and (5) [10][13]. Then, the components of the LSTM units, at a time t are updated by (1), where x_t is the input sequence, h_{t-1} is the previous block output, C_{t-1} is the previous LSTM block memory, and b_f is the polarization vector; W represents separate weight vectors for each input, and σ is the logistic sigmoid function.

$$f_t = \sigma (W_f \times [x_t,h_{t-1},C_{t-1}]+b_f) \tag{1}$$

$$i_t = \sigma (W_i \times [x_t,h_{t-1},C_{t-1}]+b_i) \tag{2}$$

$$C_t = f_t \times C_{t-1} + i_t \times tanh\ (W_C \times [x_t, h_{t-1},C_{t-1}]+b_C) \tag{3}$$

$$o_t = \sigma (W_o \times [x_t,h_{t-1},C_t]+b_o) \tag{4}$$

$$h_t = o_t \times tanh(C_t) \tag{5}$$

On the other hand, the input gate is divided into two networks and its function is to generate the new memory, as shown in Figure 1. First, a single-layer neural network takes the same inputs as the previous gate, which has a sigmoid activation function. In this part of the network, it is decided what percentage of the new memory will be influenced by the memory of the previous block. Second, a single-layer neural network, with an activation function *tanh*, generates a new memory from the input vectors (x_t) and the output from the previous block (h_{t-1}). Finally, a multiplication of vectors is performed with the outputs of the two simple neural networks, and the result is added to the output of the forgetting gate. This process is performed by equations 2 and 3. Finally, the output gate is calculated by means of equations 4 and 5, generating the probability value of the current LSTM block. In summary, the input to the cells is multiplied by the activation of the input gate, the output to the network is multiplied by that of the output gate, and the previous cell values are multiplied by the forget gate.

III. Methodology

The prediction of the class label is done by implementing a set of stages, which can be grouped into pre-processing, categorization of the data, and predictive model, as shown in Figure 2. The proposed methodology consists of three phases: data pre-processing, categorization, and construction of the predictive model. These phases allow the prediction of the label of a class based on a model of an LSTM neural network

A. Phase 1: Data Pre-processing

The data pre-processing phase consists of the following stages:

1) Data Extraction: In the data extraction stage, several activities are carried out sequentially. First, a set of attributes are extracted from the original data set. This is performed through the execution of a feature selection algorithm proposed in “in press” [14], generating a new data

 ISBN: 978-1-61208-730-6

group that will be used for the training of the neural network. Then, in the training data set, a normalization task is performed on the attribute that represents the class. Subsequently, a cleaning process is executed to detect and remove the corrupt or inaccurate instances in the database. Moreover, the attributes that contain values of a numeric type are updated, adding the first character of the name of the attribute to the value of the variable.

2) Segmentation: The segmentation task is applied to the data set generated in the previous stage. This task consists of creating a list of all the features of an instance and implementing a criterion of separation between features. Each feature is represented as a unique integer, converting the instances into a sequence of integers, thereby generating two lists of the sequence of integers. The former consists of input features (X), and the later of the output feature (Y), i.e., the class. Then, the input feature (X) is transformed into a two-dimensional matrix (the number of sequences and the maximum length of sequences).

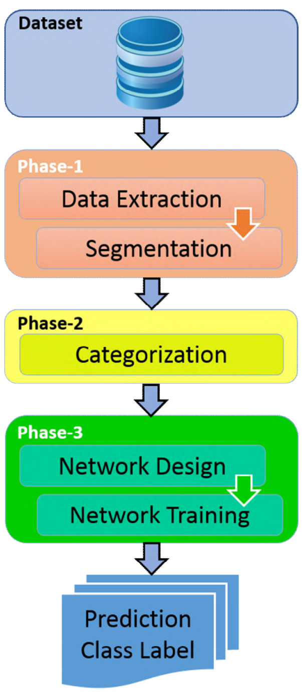

Figure 2. Methodology implemented for class prediction.

B. Phase 2: Categorization

The intermediate categorization phase consists of a process to categorize the sequence of integers corresponding to the output activities (Y), in a one-hot encoding representation type, specifying that the number of classes will be equal to the size of the vocabulary.

C. Phase 3: Prediction Model

The prediction model phase based on LSTM network is composed of the following stages:

1) Network Design: In our approach, the design of the neural network consists of generating a set of three layers. The input layer is created based on a word embedding method. The hidden layer contains the memory cells, where each cell is an LSTM unit. This layer contains a set of LSTM cells that are composed of the input, output and forget gates, which allow their interconnection. In the output layer, only one neuron will be available since there is a unique output value corresponding to the prediction. This is shown in Figure 1, where x_t is the input vector and h_t is the output result to the memory cell at time *t*. Furthermore, h_t is the value of the memory cell. At time *t*, it, ft and ot are values of the input gate, the forget gate and the output gate, respectively. Finally, C_t are values of the candidate state of the memory cell at time *t*.

2) Network Training: The training of the LSTM network is performed using as training data, the integer sequence list represented by the features contained in the two-dimensional matrix (X) and the vector that contain the class (Y) through of a representation one-hot encoding. We use the configuration parameters presented in Table I.

D. Prediction Class Label

The prediction is the output generated by the LSTM neural network, which, through a training phase, allows predicting the class label for a new instance, from a set of input features of an instance, which is explained in the following section.

TABLE I. CONFIGURATION PARAMETERS OF THE LSTM NETWORK.

Parameter	***Value***
epochs	500
batch size	32
optimizer	Adam
loss	categorical_crossentropy
LSTM	units 50

IV. EXPERIMENTAL RESULTS

Crimes are common social problems that affect the quality of life, economic growth and reputation of a country [15]. In smart cities, the aim is to reduce crime rates using ICT. Through different information systems, data is collected automatically, with the intention of generating knowledge that allows decisions to be made to reduce the criminal index. In this sense, our proposal aims to predict the value of a class, from a set of attributes. The class prediction represents the probability that an individual will be arrested after committing a crime.

The data set used is of type open data and contains historical information about the crimes to evaluate the performance of the proposed LSTM model. This data set of the reported incidents of crime that occurred in the City of Chicago from 2001 to 2018. The data set is composed of 22 attributes and 6.9 million records or instances. The complete data set can be found in [16]. Some of these attributes were

 ISBN: 978-1-61208-730-6

generated automatically by IoT devices, such as x-coordinate, y-coordinate, latitude, longitude, location. The attributes can be of data type: string, numeric, date, location or Boolean. The data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. The data set contains data from arrests, including data on Illinois Uniform Crime Reporting (IUCR), location description, domestic, date, ward, Federal Bureau of Investigation (FBI) code and more data fields. An extract from the data set is shown in Table II.

TABLE II. EXCERPT OF THE INCIDENTS OF CRIME DATA.

ID	*UICR*	*District*	*Ward*	*Com. Area*	*FBI Code*
50	479	001d	4w	32c	04B
51	820	019d	44w	6c	6
52	820	008d	18w	70c	6
53	326	006d	6w	69c	3
54	031A	025d	31w	19c	3
55	041A	007d	20w	68c	04B
56	1752	010d	12w	30c	17
57	1750	003d	5w	43c	08B
58	630	019d	2w	7c	5
59	553	022d	19w	74c	04A
60	650	010d	24w	29c	5
61	1320	009d	15w	61c	14
62	550	004d	7w	43c	04A
63	1562	006d	21w	71c	17
64	041A	017d	35w	14c	04B
65	1130	018d	42w	8c	11
66	5001	022d	19w	72c	26
67	1195	025d	29w	25c	11
68	1154	008d	13w	62c	11
69	1153	018d	43w	7c	11
70	620	018d	42w	8c	5

The feature selected to predict the label class are: UICR (Illinois Uniform Crime Reporting), District, Ward, Com. Area, and FBI Code. Table III shows a complete description of the attributes and the number of cases that were identified within the data set. The class to predict is a Boolean type, that is, *arrest = false* or *arrest = true*.

TABLE III. DESCRIPTION OF THE FEATURES SELECTED OF THE DATA SET.

Feature	*Description*	*# cases*
UICR	Illinois Uniform Crime Reporting (IUCR) codes are four-digit codes that law enforcement agencies use to classify criminal incidents.	350
District	Indicates the police district where the incident occurred.	25
Ward	The ward (City Council district) where the incident occurred.	50
Com. Area	Indicates the community area where the incident occurred.	77
FBI Code	Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS).	26
Arrest	Binary variable that indicates whether a criminal was arrested.	2

In our experiment, training data sets with 80% of the observations of the original data set are used to train our model. Instances of the training data set were selected by a random method, automatically. The remaining 20% of the instances are used to test our model prediction accuracy.

Table IV presents an extract of the results obtained in the prediction of the LSTM network using the crime data set presented before. In the column "Input Features", it is mentioned the group of the features used as a new input in the LSTM network for the prediction stage. The values of the "Input Features" column are formed by the IUCR code, district and ward id's, as well as the identifier of the community area and the FBI code. The "Target Class" is the expected label for the corresponding input instance, that is, the class label with the highest probability of prediction by the neural network, based on the weights of each class label. Each row in the table shows a case of prediction of the class from the input one. The "Output Class" column presents the label or value that the LSTM neural network predicted from the input instance for the "arrest" class.

In the column "% of prediction" (see Table IV), the probability of prediction for each instance is shown. It can be observed that, in some cases, the neural network fails to predict the class (for example 3, 7, and 18). However, in most cases, the neural network has a correct prediction of the value of the expected class. The global percentage of the prediction accuracy of the LSTM network for the complete data set is 0.87.

TABLE IV. AN EXTRACT OF THE PREDICTION FROM THE LSTM.

#	*Input Features*	*Target Class*	*Output Class*	*% of prediction*
1	0265, 011d, 28w, 29c, 02	false	false	0.82
2	0620, 025d, 31w, 22c, 05	false	false	0.97
3	0560, 005d, 34w, 53c, 08A	true	false	0.77
4	0430, 018d, 2w, 8c, 04B	false	false	0.90
5	0312, 010d, 28w, 31c, 03	false	false	0.99
6	1754, 008d, 18w, 66c, 02	true	true	0.50
7	0860, 018d, 27w, 8c, 06	true	false	0.53
8	0281, 018d, 2w, 8c, 02	false	false	0.99
9	1130, 012d, 11w, 28c, 11	false	false	0.99
10	2825, 006d, 8w, 44c, 26	false	false	0.99
11	0890, 001d, 3w, 35c, 06	false	false	0.99
12	1153, 006d, 21w, 71c, 11	false	false	0.99
13	0910, 011d, 28w, 27c, 07	false	false	0.91
14	1150, 008d, 23w, 56c, 11	false	false	0.99
15	1120, 001d, 42w, 32c, 10	false	false	0.99
16	0890, 018d, 2w, 8c, 06	false	false	0.99
17	1150, 001d, 4w, 32c, 11	false	false	0.97
18	1330, 005d, 9w, 53c, 26	false	true	0.79
19	0281, 003d, 20w, 69c, 02	false	false	0.99
20	1153, 025d, 37w, 23c, 11	false	false	0.99
21	0560, 010d, 24w, 29c, 08A	false	false	0.84
22	0820, 024d, 49w, 1c, 06	false	false	0.95
23	0110, 015d, 28w, 25c, 01A	false	false	0.88
24	1150, 019d, 32w, 6c, 11	false	false	0.99
25	1130, 020d, 48w, 77c, 11	false	false	0.99
26	0486, 011d, 28w, 26c, 08B	true	false	0.77
27	0454, 004d, 8w, 46c, 08B	false	false	0.55
28	1752, 009d, 15w, 61c, 17	false	false	0.99
29	0454, 004d, 7w, 51c, 08B	true	true	0.99
30	0454, 025d, 37w, 25c, 08B	true	true	0.97
31	0281, 008d, 22w, 56c, 02	true	false	0.65
32	0910, 012d, 25w, 28c, 07	false	false	0.82

 ISBN: 978-1-61208-730-6

33	1562, 003d, 7w, 43c, 17	true	true	0.97
34	1562, 025d, 36w, 19c, 17	true	false	0.86
35	0860, 001d, 25w, 28c, 06	false	false	0.61
36	0484, 012d, 27w, 28c, 08B	false	false	0.54
37	0430, 008d, 17w, 66c, 04B	false	false	0.94
38	1477, 014d, 26w, 24c, 15	true	true	0.77
39	1752, 007d, 16w, 67c, 17	false	false	0.54
40	0325, 025d, 36w, 19c, 03	false	false	0.99
41	0610, 025d, 31w, 19c, 05	true	false	0.82
42	0281, 008d, 17w, 66c, 02	false	false	0.98
43	1752, 006d, 17w, 44c, 17	false	false	0.76
44	0281, 018d, 42w, 8c, 02	false	false	0.99
45	1153, 018d, 42w, 8c, 11	false	false	0.95
46	0820, 016d, 41w, 9c, 06	false	false	0.98
47	1130, 012d, 28w, 28c, 11	false	false	0.99
48	1153, 008d, 14w, 63c, 11	false	false	0.99
49	1750, 022d, 34w, 75c, 08B	false	false	0.98
50	0820, 019d, 44w, 6c, 06	true	false	0.90

The LSTM network is trained using an optimization process that requires a loss function to calculate the model error. The loss function allows to faithfully summarize all aspects of the model down into a single number in such a way that improvements in that number depict a better model. Figure 3 visualizes the loss function and Figure 4 presents the accuracy of the training and validation data for the final model. The final LSTM model achieves an average loss function of 0.0376 on the validation data and a validation predictive accuracy of 0.87. Figure 4 shows the learning curves, exhibiting how the model learned and suitably fit the training data set. These results align with the loss function we obtained on the testing data, which means our final model generalizes well on new data.

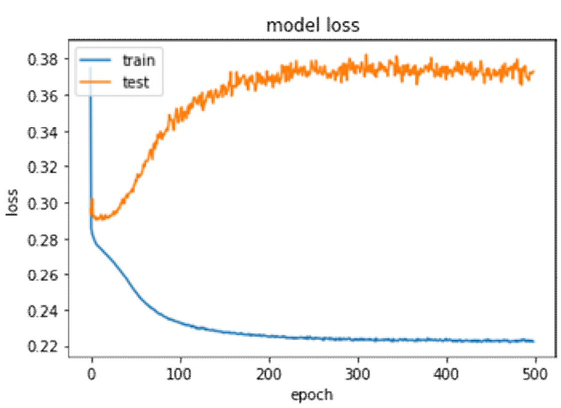

Figure 3. Loss function in training and validation phases by epoch.

V. RELATED WORK

Crime as a problem has both spatial and temporal dimensions. The work by Catlett et al. [17] explores the use of spatial analysis and auto-regressive models to automatically detect high-risk crime regions in urban areas and to reliably forecast crime trends in each region. The algorithm result is a spatio-temporal crime forecasting model, composed of a set of crime-dense regions with associated crime predictors, each one representing a predictive model for estimating the number of crimes likely to occur in its associated region. Also, in [18], the authors propose the use of deep learning for the prediction of hourly crime rates. In particular, their model is used in spatio-temporal problems, like crime forecasting. They detected that future crime rates depend on the trend set in the previous week, the time of day and the nearby events, both in space and in time. For each one of these contributing factors, they use a separate model that offers a prediction based on indicators of a weekly period, a daily period and an hourly period, respectively. The outputs of the 3 models are combined to form a common prediction.

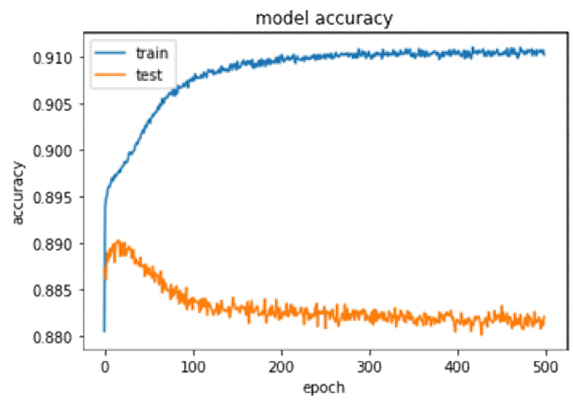

Figure 4. Accuracy in training and validation phases by epoch.

In [19], a system for the visual analysis of multidimensional data on the macroscopic and microscopic levels to show trajectories of crime based on their spatial and temporal characteristics is presented. The system incorporates a novel algorithm for the crime trajectory segmentation and uses LSTM network, generating trajectories from heterogeneous data sources, such as open data and social media, with the aim to report incidents of crime. In [20], a neural network structure for crime prediction and the appropriate inputs for crime prediction are performed through Gram-Schmidt orthogonalization for the selection of network inputs and Virtual Leave-One-Out test (VLOO) for the selection of the optimal number of hidden neurons. Spatio-temporal distribution of the hot-spots is conducted and a methodology is developed for short-term crime forecasting using the LSTM.

In [21], a context-aware attention framework is presented based on LSTM to accurately predict the amount of unrest events news. The social event prediction model consists of three parts: 1) the LSTM encoder, to get the hidden representation of the input sequence, 2) the attention layer, to automatically learn the weights of the hidden representations, and 3) the context-aware fully connected layer used to combine near historical target data and the weighted representation vectors. In [22], a crime prediction framework based on deep neural network that uncovers dynamic crime patterns and carefully explores the evolving inter-dependencies between crimes and other ubiquitous data in urban space was developed. The model enables predicting crime occurrences of different categories in each

 ISBN: 978-1-61208-730-6

region of a city, embedding all spatial, temporal, and categorical signals into hidden representation vectors.

VI. CONCLUSION

The advances in Information and Communications Technologies and the excessive use of smart devices have produced a massive generation of data. This way, IoT technologies and their implementation on the smart cities have caused a growing need for data analysis and big data analytics.

This paper investigated the effectiveness of LSTM neural networks based deep learning approach for predicting the future class labels of a crime incidents instance. To evaluate the predictive performance of our method, we use a data set that collects the historical information of crime indices in a smart city. These data are generated, in most cases, by IoT devices and managed by an information system of the police department. Before applying the LSTM model, we used a pre-processing and categorization of the data. Next, we designed and trained a neural network model in order to select the model that presents the best performance. This process is described by the proposed methodology, which guides the process of preparing input data for the LSTM neural network and its implementation.

From the original data set, 80% of the instances were used for the training stage of our model, and the remaining 20% for the validation stage of the model. The instances used in the training stage were selected by a random method automatically. Our deep learning approach achieves high performance in the final model with 87.84% accuracy based on the validation data. Furthermore, the final LSTM model achieves an average loss function of 0.0376 on validation data, using 20% of the data set for the testing stage.

The research objective was to examine the feasibility and impact of applying the proposed approach to class prediction. The experimental results suggest that the proposed method achieves good results on a big data set.

ACKNOWLEDGMENT

This work was supported by the National Council of Science and Technology (CONACYT) and Basic Science Research Project SEP-CONACYT of Mexico under Grant 256922.

REFERENCES

[1] C. Lim, K.-J. Kim, and P. P. Maglio, “Smart cities with big data: Reference models, challenges, and considerations,” Cities, vol. 82, 2018, pp. 86 – 99. [Online] https://doi.org/10.1016/j.cities.2018.04.011 [retrieved: June, 2019].

[2] E. Al Nuaimi, H. Al Neyadi, N. Mohamed, and J. Al-Jaroodi, “Applications of big data to smart cities,” Journal of Internet Services and Applications, vol. 6, no. 1, Dec 2015, p. 25. [Online] https://doi.org/10.1186/s13174-015-0041-5 [retrieved: June, 2019].

[3] C. M. Sosa-Reyna, E. Tello-Leal, and D. Lara-Alabazares, “Methodology for the model-driven development of service oriented iot applications,” Journal of Systems Architecture, vol. 90, 2018, pp. 15 – 22.

[4] I. A. T. Hashem et al., “The role of big data in smart city,” International Journal of Information Management, vol. 36, no. 5, 2016, pp. 748 – 758.

[5] S. E. Bibri and J. Krogstie, “The core enabling technologies of big data analytics and context-aware computing for smart sustainable cities: a review and synthesis,” Journal of Big Data, vol. 4, no. 1, Nov 2017, p. 38. [Online] https://doi.org/10.1186/s40537-017-0091-6 [retrieved: May, 2019].

[6] A. Rummens, W. Hardyns, and L. Pauwels, “The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context,” Applied Geography, vol. 86, 2017, pp. 255– 261.

[7] L. Vomfell, W. K. Hrdle, and S. Lessmann, “Improving crime count forecasts using twitter and taxi data,” Decision Support Systems, vol. 113, 2018, pp. 73 – 85.

[8] W. Xia et al., “Novel architecture for long short-term memory used in question classification,” Neurocomputing, vol. 299, 2018, pp. 20–31.

[9] Y. Li and H. Cao, “Prediction for tourism flow based on lstm neural network,” Procedia Computer Science, vol. 129, 2018, pp. 277–283. [Online] https://doi.org/10.1016/j.procs.2018.03.076 [retrieved: May, 2019].

[10] B. Cortez, B. Carrera, Y.-J. Kim, and J.-Y. Jung, “An architecture for emergency event prediction using lstm recurrent neural networks,” Expert Systems with Applications, vol. 97, 2018, pp. 315 – 324.

[11] C. Badii, P. Nesi, and I. Paoli, “Predicting available parking slots on critical and regular services by exploiting a range of open data,” IEEE Access, vol. 6, 2018, pp. 44059–44071. [Online] https://doi.org/10.1109/ACCESS.2018.2864157 [retrieved: May, 2019].

[12] S. Haim, S. Andrew, and B. Franoise, “Long short-term memory recurrent neural network architectures for large scale acoustic modeling,” in 15th Annual Conference of the International Speech Communication Association, September 2014, pp. 338–342.

[13] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural computation, vol. 9, no. 8, 1997, pp. 1735–1780.

[14] J.A. Mata-Torres, E. Tello-Leal, U.M. Ramirez-Alcocer, and G. Romero-Galvan, “A novel feature selection method based on a clustering algorithm,” Proc. 9th International Conference on Advances in Information Mining and Management (IMMM 2019) , in press.

[15] C. Kadar and I. Pletikosa, “Mining large-scale human mobility data for long-term crime prediction,” EPJ Data Science, vol. 7, no. 1, Jul 2018, p. 26.

[16] Chicago Police Department, “Reported Crime - Public Safety dataset,” [Online] https://data.cityofchicago.org/Public-Safety/Crimes-2001-topresent/ijzp-q8t2/data [retrieved: May, 2019].

[17] C. Catlett, E. Cesario, D. Talia, and A. Vinci, “Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments,” Pervasive and Mobile Computing, vol. 53, 2019, pp. 62-74. [Online] https://doi.org/10.1016/j.pmcj.2019.01.003 [retrieved: May, 2019].

[18] B. Wang et al., “Deep Learning for Real Time Crime Forecasting,” arXiv Numerical Analysis, pp- 1-4, 2017. [Online] arXiv:1707.03340 [retrieved: May, 2019].

[19] A. Morshed et al., “VisCrimePredict: a system for crime trajectory prediction and visualisation from heterogeneous data sources,” in Proc. 34th ACM/SIGAPP Symposium on Applied Computing (SAC19). ACM, 2019, pp. 1099-1106. https://doi.org/10.1145/3297280.3297388

[20] Z. M. Wawrzyniak et al., “Data-driven models in machine learning for crime prediction,” in Proc. 26th International Conference on Systems Engineering (ICSEng), 2018, pp. 1-8. doi: 10.1109/ICSENG.2018.8638230

[21] W. Xiuling et al., “Unrest News Amount Prediction with Context-Aware Attention LSTM”, in Proc. Trends in Artificial Intelligence (PRICAI 2018), Springer, 2018", pp. 369-377.

[22] C. Huang, J. Zhang, Y. Zheng, and N. V. Chawla, “DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction,” in Proc. 27th ACM International Conference on Information and Knowledge Management (CIKM '18). ACM, 2018, pp. 1423-1432. https://doi.org/10.1145/3269206.327179