Predict the Next Attack Location via An Attention-based Fused-SpatialTemporal LSTM

1st Zhuang Liu
Engineering Research Center of ACAT, Ministry of Education.
Beihang University
Beijing, China
liuzhuang@buaa.edu.cn

2nd Juhua Pu
Research Institute of Beihang University in Shenzhen.
Beihang University
Shenzhen, China
pujh@buaa.edu.cn

3rd Nana Zhan
Engineering Research Center of ACAT, Ministry of Education.
Beihang University
Beijing, China
zhannana@buaa.edu.cn

4th Xingwu Liu
Institute of Computing Technology.
Chinese Academy of Sciences
Beijing, China
liuxingwu@ict.ac.cn

***Abstract*—With the frequent occurrence of unconventional global emergencies, the public security field has received more and more attention. As an unconventional emergency, terrorist attacks have aroused global attention. So, how should we extract useful information from a large number of terrorist attacks and find the law of the attack, so that we can effectively prevent or take early measures to reduce losses? To this end, we are based on the Global Terrorism Database (GTD), and aim to predict the next province or state a terrorist organization may attack at a specific time point by mining the terrorist organizations' historical records and other types of information availabl, such as incident information and so on. Then, Based on these incident information and spatiotemporal information, we propose a neural network called ATtention-based Fused-SpatialTemporal LSTM (ATFST-LSTM) to predict the next location which may be attacked. We test the efficiency of our models on GTD, experiments show that our models has achieved better results.**

***Index Terms*—terrorist attack, global terrorism database, incident information, next location prediction**

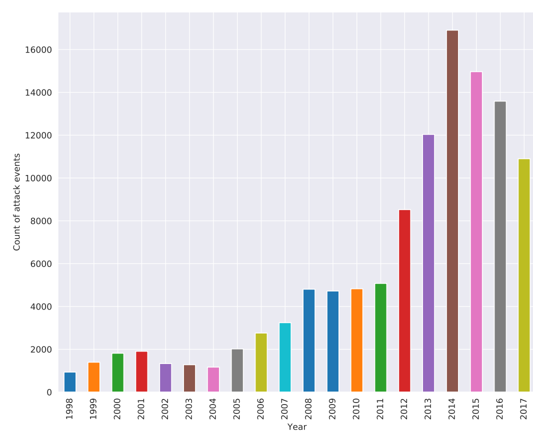

Fig. 1. The number of terrorist attacks from 1998 to 2017.

I. Introduction

Terrorism, which is a complex and long-lasting social phenomenon, has become one of the greatest threats to peace and security in the international community [1]. According to the statistics of the Global Terrorism Database (GTD), the number of terrorist attacks launched by terrorist organizations before 2005 was relatively stable. But the number of terrorist attacks since 2005 has increased by an order of magnitude. Only in 2014, 16817 terrorist attacks were launched. As is shown in Fig. 1.

These terrorist activities have caused a lot of property damage and casualties. Therefore, to help governments and police departments to effectively prevent terrorist attack from happening or to handle them efficiently when they occur, accurate and reliable prediction of terrorist attack is a necessity. For example, if we can predict the next province or state a terrorist organization may attack at a specific time point, so that we take measures in advance to minimize the losses caused by it. In this paper, we aim to solve this problem by mining the terrorist organizations' historical records and other types of information available.

Many researches have been done on analyzing terrorism incident data around the world. For example, Hawkes Process is applied to predict terrorist attacks in Northern Ireland which considered 5000 explosions ranging between year 1970-1998 to predict when and where the Irish Republicans Army (IRA) launched attack in Northern Ireland which considered 5000 explosions ranging between year 1970-1998 [2]. Faryal et al. [3] proposed a Terrorist Group Prediction Model (TGPM) to predict the group involved in a specific attack. In [4], Dynamic

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0803700, Science Technology and Innovation Commission of Shenzhen Municipality(JCYJ20180307123659504), the State Key Laboratory of Software Development Environment.

978-1-7281-6607-0/20/$31.00 ©2020 IEEE

Bayesian Network (DBN) are proposed to predict the likelihood of future attacks, which acts an initial step in predicting the terrorist behavior at critical transport infrastructure facility. In addtion, Social Network Analysis (SNA) is proposed to predict whether a person is terrorist or not and resulted in 86% accuracy [5]. Only a small number of schemes have been proposed more recently to study the next location prediction by exploring the spatial and temporal patterns of terrorist attacks [6].

These works at present are good attempts on the task of terrorist analysis and have achieved certain results. However, these solutions did not fully handle the noisy data, which includes missing values and many other redundant features for a specific tasks. Moreover, they ignore the terrorist attack incident information (such as the summary information, attack information, weapon information and so on.). These information is important to improve the results for the next location prediction task. For example, in the GTD data source, there is a "multiple" field to record whether a certain attack incident is related with other incidents. When an attack incident is part of an associated incident group, the relevant incident number is recorded. This shows that these incidents are not independent of each other, and the occurrence of one incident may lead to the occurrence of another incident. Therefore, how to fully mine the complex terrorist attack information to improve the accuracy of predicting terrorist organizations' next attack location is still a challenging issue.

To address the aforementioned problems, in this paper, we use Normalized Mutual Information (NMI) [7] to select relevant features for location, then we propose a neural network called Attention-based Fused-SpatialTemporal LSTM (ATFST-LSTM) to model the spatiotemporal contextual information and the terrorist attack incident information. In addition, we use the artificial "attention mechanism" in neural networks [8]–[10] to capture the terrorist organizations' preference different weights.

The main contributions of this work are summarized as follows:

- We propose a neural network called ATFST-LSTM based on the terrorist organizations' historical records and other types of information available to predict the next location which a terrorist organization will attack.
- Experiments conducted on real-world datasets GTD show that ATFST-LSTM is effective and outperforms the state-of-the-art methods.

II. Related Works

Terrorist attacks are a common phenomenon and are extremely harmful worldwide. With the continuous accumulation of terrorist attack data and the development of big data technology, people have begun to focus on analyzing the data and discovering the rules of the attacks, so as to conduct more targeted prevention and control.

After the "911" incident, Edelstein et al. [11], Derosa et al. [12] proposed the techniques to deal with terrorism. With the emergency of new data mining algorithms, data mining is more widely used in anti-terrorism. They discussed the collection and processing of terrorist information, the choice of models, the risk control of terrorism, and the deepening of the data mining foundation in the field of counter-terrorism.

In this section, we will introduce the related works from the flowing parts:

Terrorist behavior research. The current research mainly focuses on the prediction and assessment of terrorist behavior. For the prediction task, in 2005, Carley et al. [13] began to use dynamic network analysis techniques to study the behavior of terrorist groups. Schrodt et al. [14] used the Hidden Markov model to predict conflict in the balkansusing from 1991 to 1999. Reghavan et al. [15] used the hidden Markov model to establish a model for a terrorist organization's activity and detect the sudden situation of the organization. Ali et al. [16] used data mining technology to identify terrorist organizations from social network information. Recently, Qiang et al. [6]extended RNN and proposed a novel method called Spatial Temporal Recurrent Neural Networks (ST-RNN) to predict the next location for terrorist groups. ST-RNN can model local temporal and spatial contexts in each layer with time-specific transition matrices for different time intervals and distance-specific transition matrices for different geographical distances. But they ignore the terrorist attack incident information. In the field of risk assessment for terrorist attacks, a terrorist attack prediction project led by Blair et al. used a neural network to successfully predict the conflict in Liberia in 2010 with the data in 2008; the accuracy was between 0.65 and 0.74 [17]. In addition the machine learning-based terrorist attack assessment model can also be widely accommodated and integrated with unstructured data, and it has the ability to find discernable patterns from clutter and mixed data [18].

Next location prediction. Next location prediction, more known as next POI recommendation, is recently proposed and has attracted great research interest. Most of the existing works usually employ the properties of a Markov chain to model the sequential patterns of users. For example, a tensor-based model, named FPMC-LR, was proposed by integrating the first-order Markov chain of POI transitions and distance constraints for next POI recommendation [19]. Similarly, in [20], an additive Markov chain model for predicting the sequential transitive probability. And Ye et al. [21] proposed a mixed hidden Markov model to learn the POI categories' transitive patterns of sequential user check-ins. In addition, recurrent neural networks are used for modeling the sequential of user check-ins. For example, Yang et al. [22] employed the RNN and gated recurrent unit (GRU) models to characterize short-term and long-term sequential contexts separately. In [6], ST-RNN is proposed for location prediction, which utilized an RNN architecture to learn the sequential transition. In [23], ATST-LSTM is proposed to learn the non-linear dependency representation over POIs and the spatio-temporal contexts from historical check-in activities.

Unlike the above studies, in the terrorist attack fields, there are much more information of attack incident (such as the summary information, attack information, weapon information

TABLE I
NOTATIONS USED IN THIS PAPER.

Symbol	Description
u	terrorist organization
v	attack location
t	time for a terrorist attack
e	incident information
U	the set of terrorist organizations
$v_{t_k}^u$	location attacked by u at t_k
$\mathbf{P}_u \in \mathbb{R}^d$	latent representation of u
$\mathbf{v}_{t_k}^u$	embedding vector of location $v_{t_k}^u$
Δt_i	time interval
Δd_i	geographical distance

and so on), how to handle these noisy data well and extract effective information from them is critical. Therefore, in this paper, we propose a neural network called Attention-based Fused-SpatialTemporal LSTM (ATFST-LSTM) to model the spatiotemporal contextual information and the terrorist attack incident information. Based on them, we aim to predict the next location a terrorist organization may attack at a specific time point.

III. Methodology

In this section, we first present the problem description of our study. Then, we present the details of our ATFST-LSTM model. Finally, in order to handle the terrorist attack incidents information, we present the details of our methods.

A. Problem Description

We begin with some necessary notations (as is shown in Table I) and then formally present the problem formulation of terrorist organizations' next location prediction.

Definition 1: Location. Location is a spatial item that records the province or state of a terrorist attack incident. It associates with latitude, longitude and name.

Definition 2: Terrorist attack incident. A terrorist attack incident by a terrorist organization is a quadri-tuple $c_{t_k}^u = (u, v_{t_k}^u, e_{t_k}^u, t_k)$, which indicates that a terrorist organization u attacks location $v_{t_k}^u$ at time point t_k and $e_{t_k}^u$ represent the incident information.

Definition 3: Incident sequence. An incident sequence of a terrorist organization u is a set of terrorist attack incidents, denoted by $C_u = (c_{t_1}^u, c_{t_2}^u, \cdots, c_{t_i}^u)$. For simplicity, the historical incidents of all the terrorist organization are denoted by $C^U = (C_{u_1}, C_{u_2}, \cdots, C_{u_{|U|}})$, where U is the set of terrorist organizations.

The primary goal of this work is to offer a list of possible locations that a terrorist organization is likely to attack at the next time point by mining all terrorist organizations' attack records. Here, we formulate the problem as follows:

Definition 4: Next location prediction. Given a terrorist organizations' incident sequence $C_u = (c_{t_1}^u, c_{t_2}^u, \cdots, c_{t_{i-1}}^u)$, the goal of next location prediction is to calculate a score for each location v_k based on C_u and a time point t_i. Higher score indicates higher probability that a terrorist organization u will attack that location at time t_i.

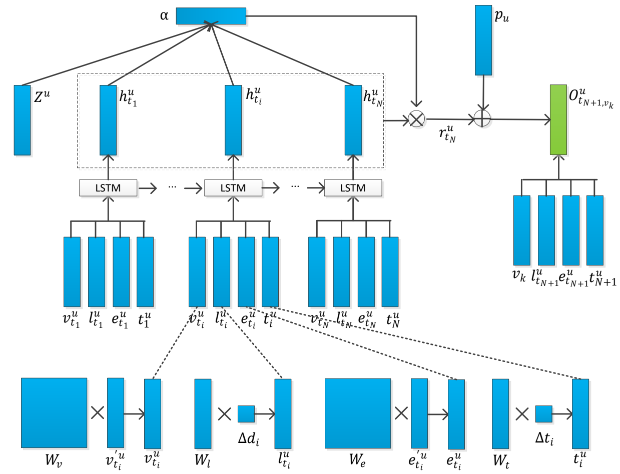

Fig. 2. The architecture of ATFST-LSTM.

B. Attention-based Fused-SpatialTemporal LSTM

As we all known, spatio-temporal and incident information have a great impact on terrorist attacks, it is essential to model these information to predict the next location of a terrorist organization. Therfore, we propose a neural network called Attention-based Fused-SpatioTemporal LSTM (ATFST-LSTM).

Fig. 2 illustrates the architecture of ATFST-LSTM. Our model is based on LSTM network. In order to model the spatio-temporal information better, we consider both time intervals and geographical distances between neighbor items, which is helpful to the next location prediction [24]. At time point t_i, we define $\Delta t_i = t_i - t_{i-1}$ and $\Delta d_i = v_i - v_{i-1}$ represent the time intervals and geographical distances, respectively. For a terrorist organization u, we define the location embedding and incident embedding as $\mathbf{v}_{t_i}^u$ and $\mathbf{e}_{t_i}^u$ (we will discuss how to obtain the incident embedding $e_{t_i}^u$ in Sect. III-C) at time point t_i. And we can obtain a quadri-tuple $(\mathbf{v}_{t_i}^u, \Delta d_i, \mathbf{e}_{t_i}^u, \Delta t_i)$. Then, ATFST-LSTM takes the quadri-tuple as the input at each time step.

After receiving the current input and the memory $\mathbf{h}_{t_{i-1}}^u$ from the past information, ATFST-LSTM performs a linear transformation on the quadruple vectors separately, and then input the transformed vector into the LSTM unit. Finally, ATFST-LSTM gets the hidden vector $\mathbf{h}_{t_i}^u$ at time t_i. More formally,

$$\mathbf{h}_{t_i}^u = LSTM(\mathbf{W}_v \mathbf{v}_{t_i}^u + \mathbf{W}_l \Delta d_i + \mathbf{W}_e \mathbf{e}_{t_i}^u + \mathbf{W}_t \Delta t_i, \mathbf{h}_{t_{i-1}}^u), \quad (1)$$

where $\mathbf{W}_v \in \mathbb{R}^{d \times d}$, $\mathbf{W}_l \in \mathbb{R}^{d \times 1}$, $\mathbf{W}_e \in \mathbb{R}^{d \times d}$, $\mathbf{W}_t \in \mathbb{R}^{d \times 1}$ are transition matrices, d is the dimension of hidden layer. Here, $h_{t_i}^u$ is dynamically changing with the time point, and can be regarded as the terrorist organization u's preferences for locations under different spatial-temporal and event contexts.

Generally speaking, not all historical information are related equally to a terrorist organization u's next location. Therefore, we design an attention mechanism after gaining the hidden layer's output $\mathbf{h}^u_{t_i}$. By using the attention mechanism, ATFST-LSTM can help to assign the terrorist organization u's preference different weights at different time points.

Here, we define a parameter $\mathbf{Z}^u$ as the query of our attention model. Then, we calculate the dot-product between each time-step's output $\mathbf{h}^u_{t_i}$ and $\mathbf{Z}^u$. To prevent the result from being too larger when the dimension d of hidden layer is bigger, we define the attention function with a scale:

$$f(\mathbf{h}^u_{t_i}, \mathbf{Z}^u) = \frac{\mathbf{h}^u_{t_i}(\mathbf{Z}^u)^T}{\sqrt{d}} \tag{2}$$

Then we use the softmax function to normalize the result to get the attention weight α_i at each time point t_i, namely,

$$\alpha_i = \frac{exp(f(\mathbf{h}^u_{t_i}, \mathbf{Z}^u))}{\sum_{i=1}^{N} exp(f(\mathbf{h}^u_{t_i}, \mathbf{Z}^u))}, \tag{3}$$

where N denotes the length of the input sequence. By using the attention mechanism, we mixed each attention weight α_i and the corresponding hidden vector of the input sequence to produce a weighted hidden vector

$$\mathbf{r}^u_{t_N} = \sum_{i=1}^{N} \alpha_i \mathbf{h}^u_{t_i} \tag{4}$$

In addition, in order to represent the static information of each terrorist organization u, we define an embedding vector p_u for each u. Thus, every terrorist organization u is represented by dynamic mode $h^u_{t_i}$ and static mode p_u. Finally, the predicted probability that terrorist organization u go to location v_k at time point t_{N+1} is calculated as the following,

$$\begin{gathered} O^u_{t_{N+1},v_k} = (\mathbf{W}_N \mathbf{r}^u_{t_N} + \mathbf{W}_p \mathbf{p}_u)^T \\ (\mathbf{W}_v \mathbf{v}_k + \mathbf{W}_l \Delta d_i + \mathbf{W}_e \mathbf{e}^u_{t_i} + \mathbf{W}_t \Delta d_i) \end{gathered} \tag{5}$$

In order to train our model, we use the Bayesian Personalized Ranking (BPR) and Back Propagation Through Time (BPTT). The basic assumption of BPR is that a user prefers a selected location than a negative one. Then, we need to maximize the following probability:

$$p(u, t, v \succ v') = g(O^u_{t,v} - O^u_{t,v'}), \tag{6}$$

where v' denotes a negative location sample, and $g(x)$ is a nonlinear function which is selected as: $g(x) = \frac{1}{1+e^{-x}}$.

Incorporating the negative log likelihood, we can solve the following objective function equivalently:

$$J = \sum ln(1 + e^{-g(O^u_{t,v} - O^u_{t,v'})}) + \frac{\lambda}{2}||\theta||^2, \tag{7}$$

where λ is used to determine the power of regularization and θ is the parameter set. Besides, we use AdaGrad to optimize the network parameters in this study.

C. Incident Information Extraction

In this subsection, we will present the detail of how to obtain the incident embedding mentioned in Sect. III-B.

The data set of GTD is a noisy data, which includes unknown values, missing values. And it contains up to 135 features. In order to select relevant features that contributes in predicting the next location for a terrorist organization, we use NMI to measure the correlation between each feature and location. The larger NMI value of the feature and location, the stronger correlation between them. We select the more relevant features as incident information. The selected features include categorical features (such as attack information, weapon information), numerical features (such as casualties, property damage) and textual features (such as incident summary). In order to obtain the incident embedding **e** better, we first process them by different methods.

In the GTD data source, most of the fields are categorical features. For example, the "city" field has 25853 categories. Here, we make embedding representations of these features. We map each categorical feature in a function approximation problem into Euclidean spaces, which are the entity embeddings of the categorical variable. For numerical features, we analyze them and calculate the skewness and kurtosis respectively. Here, taking the number of deaths as an example, the skewness value is 58.51, the kurtosis value is 5991.48. It can be found that the values of skewness and the kurtosis are large and seriously deviates from the normal distribution. Therefore, it is necessary to perform log conversion on the target value to restore the normality of the target value. For textual features, in order to extract the keywords, we use the TF-IDF weighting technique [25]. Here, we take the summary field of each incident in the GTD data source as a document, and the summary information of all events as the entire corpus.

Then, we concatenate all the above vectors as the incident embedding **e**. Besides, the dimension of **e** may not be consistent with the embedding dimension of the location, to make the formula (1) operable, we use Singular Value Decomposition (SVD) to reduce the dimension of **e** to d. The incident information vector **e** can be continuously iteratively updated in the model of ATFST-LSTM.

IV. EXPERIMENTS

In our experiments, we verify the prediction effect of the ATFST-LSTM model with and without the attention mechanism, respectively. (Here, to verify the effect of incident information on predicting the next location, we removed the attention mechanism in the ATFST-LSTM model, and defined this model as FST-LSTM). In addition, We perform experiments to evaluate the performance of FST-LSTM, ATFST-LSTM on the real-world datasets GTD. In particular, we aim to answer the following questions:

Q1: How does our FST-LSTM model perform as compared to the baseline methods such as ST-LSTM [24] and ATST-LSTM [23]?

Q2: Does ATFST-LSTM improve the performance of FST-LSTM by using attention mechanism?

A. Data Statistics

The data set for this study is used from https://www.start.umd.edu/gtd/ i.e. GTD, which has 135 attributes, numbered 1-135. For each attribute, its completeness is calculated separately. That is, the completeness of a attribute is the ratio of

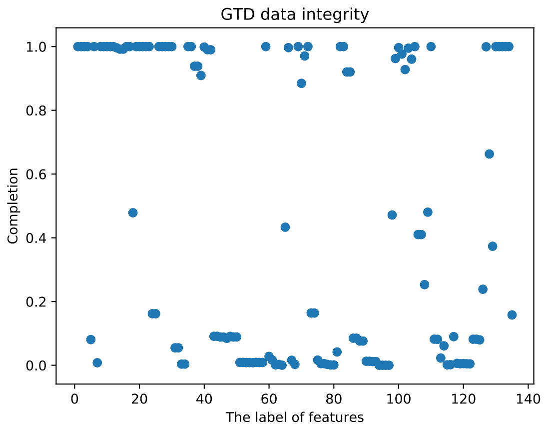

Fig. 3. Data completeness in GTD database.

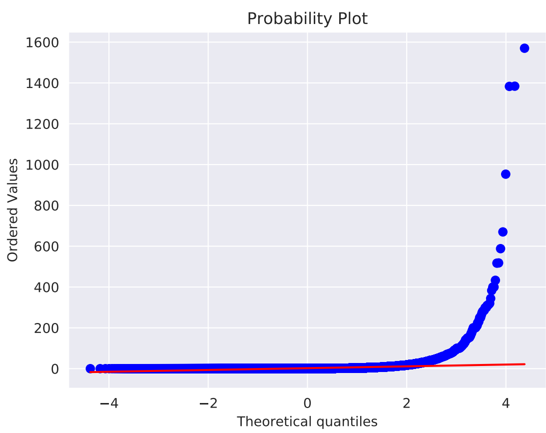

Fig. 4. Distribution of deaths

the number of samples without missing values in this attribute to the total number of samples. The result is shown in Fig. 3.

As can be seen from Fig. 3, there are 57 attributes with a completeness of 80% or more, 68 with a completeness of 20% or less, and the rest are between 20% and 80%. Here, we only analyze the 57 attributes with higher integrity and discard the remaining features. Among the selected features, some are different representations of the same feature. For example, the "country" feature is represented by both a numerical and a country name. For this type of features, we keep only one way. For numerical features, such as the number of deaths of each incident, a normal probability map is drawn, as shown in Fig. 4. For missing values, we use a median fill for numerical features and mode fill for categorical features.

B. Evaluations on ATFST-LSTM

In order to validate the effectiveness of our proposed model ATFST-LSTM, we compare it with FST-LSTM and following methods for next location prediction.

TABLE II
DISCRETE AND CONTINUOUS FEATURES WITH NMI VALUE GREATER THAN 0.05 OF LOCATION.

features	NMI
latitude	0.8063
longitude	0.8061
victim nationality	0.6939
organization name	0.4935
year	0.2247
target type	0.1057
attack type	0.0806
weapon type	0.0770

- LSTM [26]: This is a variant of RNN model, which contains a memory cell and three multiplicative gates to allow long-term dependency learning.
- ST-LSTM [24]: It implements time gates and distance gates into LSTM to capture the spatio-temporal relation between successive check-ins. And it utilizes time and distance intervals to model user's short-term interest and long-term interest simultaneously.
- ATST-LSTM [23]: This is an attention-based spatio-temporal LSTM, which is to learn the non-linear dependency representation over POIs and the spatio-temporal contexts from historical check-in activities.

In this experiment, we remove all the terrorist organizations with fewer than 10 occurrences in GTD and all the locations where were attacked with less than 10 occurrences. Next, we take the remaining records in time order for each terrorist organization. Finally, 80% records of the behavioral history of each terrorist organization are selected for training, 20% for testing. In the training set, NMI is used to select the categorical features and numerical features associated with the location. Here we take the feature set with the NMI value greater than 0.05, as is shown in Table II. And we select incident summary information as the textual features. These features are regarded as the incident information, which are the input for MFNN. In addition, The regulation parameter for this experiment is set as $\lambda = 0.01$.

We employ three evaluation metrics **Precision@k**, **Recall@k** and **F1-score@k** for ranking tasks. The evaluation score for our experiment is computed according to where the next selected location appears in the ranked list. We report precision@k, recall@k and F1-score@k with k=1,5 and 10 in our experiments. The larger the value, the better the performance. Besides, we employ Mean Average Precision (**MAP**) for global evaluations for the quality of the whole ranked lists.

The performance comparison on GTD is illustrated in Table III. Here, FST-LSTM is a model without attention mechanism. Compared with the LSTM and ST-LSTM models without attention mechanism in the same, it has improved to different degrees on different metrics, which indicates the terrorist attack incident information has a positive effect for predicting the next location. In addition, the ATFST-LSTM model with the attention mechanism has an improvement

TABLE III
THE RESULTS OF NEXT LOCATION PREDICTION.

Algorithm	P@1	P@5	P@10	R@1	R@5	R@10	F1@1	F1@5	F1@10	MAP
LSTM	0.2865	0.0601	0.0354	0.2865	0.3469	0.4225	0.2865	0.1103	0.0659	0.3163
ST-LSTM	0.3019	0.0738	0.0431	0.3019	0.3692	0.4308	0.3019	0.1231	0.0783	0.3384
ATST-LSTM	0.3046	0.0765	0.0433	0.3046	0.3827	0.4327	0.3046	0.1276	0.0787	0.3464
FST-LSTM	0.3038	0.0769	0.0450	0.3038	0.3846	0.4503	0.3038	0.1282	0.0818	0.3451
ATFST-LSTM	**0.3192**	**0.0781**	**0.0454**	**0.3192**	**0.3904**	**0.4538**	**0.3192**	**0.1301**	**0.0825**	**0.3533**

of about 1 percentage point compared with the FST-LSTM, which shows the effectiveness of the attention mechanism. Similarly, our proposed ATFST-LSTM model is also improved compared to the ATST-LSTM model. This is the joint contribution of terrorist attack incident information and attention mechanism to the prediction results. From this experimental results, we can see, the information about the terrorist attack incident is significant on modeling terrorist organizations' behavior.

V. CONCLUSIONS

In this paper, we propose a neural network called ATFST-LSTM to predict the next location for terrorist organizations, which incorporates the spatio-temporal and incident information available on terrorist attack. And by the attention mechanism, ATFST-LSTM can capture the terrorist organizations' preference different weights. Experiments show that ATFST-LSTM outperform the baseline methods in next location prediction task, it is indicated that the incident information is important to predict the terrorist organizations' behavior. In the future work, we plan to combine much more features to do some multivariate prediction applications.

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for their valuable comments and helpful suggestions. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0803700, Science Technology and Innovation Commission of Shenzhen Municipality(JCYJ20180307123659504), the State Key Laboratory of Software Development Environment.

REFERENCES

[1] U. Beck, "The terrorist threat: World risk society revisited," *Theory, culture & society*, vol. 19, no. 4, pp. 39–55, 2002.
[2] S. Ana, "The eerie math that could predict terrorist attacks," *International Journal of the Care of the Injured*, vol. 47, no. 3, pp. 646–652, 2016.
[3] F. Gohar, W. H. Butt, U. Qamar *et al.*, "Terrorist group prediction using data classification," *Work. MultiRelational Data Min. MRDM2003*, vol. 10, pp. 199–208, 2014.
[4] M. K. Jha, "Dynamic bayesian network for predicting the likelihood of a terrorist attack at critical transportation infrastructure facilities," *Journal of Infrastructure Systems*, vol. 15, no. 1, pp. 31–39, 2009.
[5] T. R. Coffman and S. E. Marcus, "Pattern classification in social network analysis: A case study," in *2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No. 04TH8720)*, vol. 5. IEEE, 2004, pp. 3162–3175.
[6] L. Qiang, W. Shu, W. Liang, and T. Tan, "Predicting the next location: a recurrent model with spatial and temporal contexts," in *Thirtieth Aaai Conference on Artificial Intelligence*, 2016.
[7] P. A. Estevez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 189–201, 2009.
[8] V. Mnih, N. Heess, A. Graves *et al.*, "Recurrent models of visual attention," in *Advances in neural information processing systems*, 2014, pp. 2204–2212.
[9] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
[11] H. Edelstein, "Using data mining to find terrorists," *Data Mining Review*, 2003.
[12] M. DeRosa, *Data mining and data analysis for counterterrorism*. CSIS Press Washington, DC, 2004.
[13] K. M. Carley, "Dynamic network analysis for counter-terrorism," *Unpublished manuscript*, 2005.
[14] P. A. Schrodt, "Forecasting conflict in the balkans using hidden markov models," in *Programming for Peace*. Springer, 2006, pp. 161–184.
[15] V. Raghavan and A. G. Tartakovsky, "Hidden markov models for the activity profile of terrorist groups," *Annals of Applied Statistics*, vol. 7, no. 4, pp. 2402–2430, 2013.
[16] G. A. Ali, "Identifying terrorist affiliations through social network analysis using data mining techniques," 2016.
[17] A. Hartman, R. Blair, and C. Blattman, "Predicting local violence: Evidence from a panel survey in liberia," *Social Science Electronic Publishing*, vol. 54, no. 2, 2017.
[18] Q. Dong, "Machine learning and conflict prediction:a cross-disciplinary approach," *World Economics & Politics*, 2017.
[19] C. Cheng, H. Yang, M. R. Lyu, and I. King, "Where you like to go next: Successive point-of-interest recommendation," in *Twenty-Third international joint conference on Artificial Intelligence*, 2013.
[20] J.-D. Zhang, C.-Y. Chow, and Y. Li, "Lore: Exploiting sequential influence for location recommendations," in *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2014, pp. 103–112.
[21] J. Ye, Z. Zhu, and H. Cheng, "What's your next move: User activity prediction in location-based social networks," in *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM, 2013, pp. 171–179.
[22] C. Yang, M. Sun, W. X. Zhao, Z. Liu, and E. Y. Chang, "A neural network approach to jointly modeling social networks and mobile trajectories," *ACM Transactions on Information Systems (TOIS)*, vol. 35, no. 4, p. 36, 2017.
[23] L. Huang, Y. Ma, S. Wang, and Y. Liu, "An attention-based spatiotemporal lstm network for next poi recommendation," *IEEE Transactions on Services Computing*, 2019.
[24] P. Zhao, H. Zhu, Y. Liu, Z. Li, J. Xu, and V. S. Sheng, "Where to go next: A spatio-temporal lstm model for next poi recommendation," *arXiv preprint arXiv:1806.06671*, 2018.
[25] H. C. Wu, R. W. P. Luk, K. F. Wong, and K. L. Kwok, "Interpreting tf-idf term weights as making relevance decisions," *Acm Transactions on Information Systems*, vol. 26, no. 3, pp. 55–59, 2008.
[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.