Traffic Flow Prediction and Application of Smart City Based on Industry 4.0 and Big Data Analysis

Wiam El Bouchti

December 15, 2022

Traffic Flow Prediction and Application of Smart City Based on Industry 4.0 and Big Data Analysis

1st WIAM ELBOUCHTI
IPSS student
faculty of sciences Rabat
wiam.elbouchti@um5r.ac.ma

***Abstract*—The prediction of traffic accidents is important for improving transportation safety as well as route safety. The problem is also difficult because the causes are complex and differing such as mechanical problems of the vehicle, negligence of the driver and even the factors that can cause traffic accidents vary from place to place. For example, the main factors leading to traffic accidents in an urban region with busy local roads may be very different from those in a rural expressway, and accidents are rare events. It is difficult to accurately predict individual accidents due to the lack of sufficient samples.**

In this paper, we perform an in-depth study of the problem of traffic accident prediction using the Convolutional Long Short-Term Memory (ConvLSTM) neural network model. A number of detailed features such as weather, environment, road conditions and traffic volume

I. Introduction

Smart City is a city equipped with different technological solutions that collects data from its inhabitants to provide them with better information to improve the efficiency of public services. This data collection concerns all areas such as health, culture, security, housing, education, tourism, social cohesion, etc. By creating data that is accessible, shareable and usable by all, i.e. the possibility for different systems or companies to work together by mixing their data.

The case for the health field; In a more ambitious way, data can be put to good use. This ranges from an increased use of digitalization of medical monitoring to genomic sequencing to obtain fully personalized treatments. This intelligence also allows patients to have a virtual visit from their doctor, who can even send them the medication they need.

In spite of this important data collection, the data are not yet exploited in their entirety. every year, more than 1.3 million people (motorists, motorcyclists and pedestrians) lose their lives in road accidents, causing considerable economic for those who die or are disabled as a result of their injuries, as well as for the family members for the family members.

In order to reduce the number of injuries, there are several essential elements to consider eliminate fatal accidents;

- Every driver must respect the Highway Code, which gathers all the rules and laws to be respected when driving on public roads. when driving on the public road and its purpose is to ensure the safety and and the fluidity of traffic on the roads.

Identify applicable funding agency here. If none, delete this.

- The driver must be vigilant and adapt their behavior to each situation and conditions such as night, snow, rain, etc.

- Regular checks of the vehicle's equipment (fluid levels, tire pressure and wear, brake pads, etc.) and carrying out the mandatory technical checks.

- Drivers must concentrate on their steering wheel; Using a phone while driving lengthens reaction times (particularly regarding braking or road traffic signals) and makes it more difficult to keep the vehicle in the right traffic lane and to respect the safety distances with the vehicle in front .

- Avoid alcohol and all psychoactive substances that can promote drowsiness.

- It must be ensured that all provisions are made such as sidewalks, cycle paths, protected passages to cross the roadway and other devices intended to slow down traffic.

Solutions are currently being developed but remain at . They must be dynamic, open and adapted by relying on new technologies. If the development is long and tedious, everyone will be a winner quickly.

II. RELATED WORK

A. *Existing studies*

The detection of accidents is currently done in a traditional way: camera surveillance of servers , declaration..) This type of detection has many disadvantages such as: Constraint of the limited number of streets asks for the existence of the material; there will be no expectation of accidents; In order to solve these drawbacks, several tools have been created based on new technologies. Among which we can mention:

- WAZE
  Waze uses real-time data from the app's users to provide the best route to the user's destination, taking into account accidents, traffic jams, speed traps, construction, and other obstacles that could slow down a driver. Users store a list of friends on the app so that they can keep track of each other on a trip or spot a friend in the vicinity.
  Updates are submitted automatically as users drive and also can be proactively shared via the app.
- SALAMATI
  Detection and prevention of road accidents in Morocco in real time, Salamati is a prototype mobile application that automatically detects and sends SMS alerts to notify

the contacts of the damaged vehicle, specifying the geolocation. Among these objectives: - Alert Families And Friends Of Accident Victims By SMS To Intervene And Help Rapidly. - Alert the nearest local authorities of the occurrence of an accident with Localization To Reduce Intervention Time.

- Personal Safety
  Personal Safety is a Pixel app that helps you prepare and react in an emergency by quickly connecting you with the help and information you need.

B. KNOWLEDGE

At the beginning of the 1980s, China experienced the birth of economic zones which authorized Foreign investments ; among these areas there was SHENZEN village. This little village took 30 years to become China's ambitions from ecology to intelligence; SHENZEN has integrated buses and electronic taxis as well as a health service system that allows enablers to track their folders.

During these 30 years China has been able to go from 18 percent of the population living in cities to 58 percent of inhabitants, it has experienced the creation of 160 cities and it aims to have 292 million in the 2050s, for this reason the Chinese government created in 2018 19 new groupings of cities among these cities the first smart city in the world of more than 40 million inhabitants plus even the most populous city in the world SHANGHAI.

Among the most important public safety problems in China are road accidents. circulation, every year more than 1.3 million people die According to the World Health Organization health (WHO), most of these die are young people between 15 to 29 years. The reduction of road accidents are a crucial societal problem. The ability to understand and predict potential accidents in the future (where, when or how) is therefore very useful not only to the actors public safety (the police), but also to transport administrators and individual travellers.

C. problematic

Forecasting traffic accidents is a very difficult problem. First, the causes traffic accidents are complex such as vehicle mechanical problems, driver negligence Second, traffic accidents are rare occurrences. Finally, the factors that can cause traffic accidents vary from place to place. For example, the main factors that lead to traffic accidents in a region urban with busy local roads can be very different from those of a lane rural express. Managing the spatial heterogeneity of data is a challenge.

Generally, the traffic accident prediction problem has been formulated as a classification problem or a regression problem. For example, some work aims to predict whether or not an accident will occur at a specific location. Other works predict the number of accidents at a given time and place using regression models. However, these works generally use classical data mining methods and do not take into account the unique characteristics of traffic accident data such as spatial heterogeneity and temporal autocorrelation, which leads to unsatisfactory performance.

A limited number of recent works have attempted to solve the problem using Deep Learning, such as the convolutional neural network. In this study, the problem is formulated as a image prediction problem, where a traffic risk map is generated by learning from traffic accident records and other data such as traffic volume, road condition, precipitation, temperature and satellite images are collected and matched to each grid cell. Considering the number of accidents as well as other urban and environmental characteristics at each location,

III. RELATED WORK

Prediction of traffic accidents using conventional techniques

large body of literature by public safety and injury prevention researchers attempts to classify each given segment of road at any given time into Accident, No Accident binary classes. There was a comparison between the performance of the Artificial Neural Network with that of a negative binomial regression model on 1338 crashes. ANN achieved 64 PERCENT and 61 PERCENT accuracy for training and testing, respectively. Or they also applied the decision tree model on the same dataSet to predict road accidents. Training and testing accuracy is less than 55 PERCENT . Other researchers applied a decision tree and an ANN model on a Nigeria dataset and obtained precision and recall of around 0.52 percent .

All the above works simply apply classical data mining techniques on small-scale traffic accident data (one or a small number of roads) with limited features. Also, they generally do not address unique data properties such as temporal periodicity, spatial autocorrelation, and heterogeneity, thus having relatively low accuracy.

Deep Learning for Traffic Accident Prediction

Some recent work has tried to solve the problem of traffic accident analysis using Deep Learning. Some have used mobile phone data and historical crash records to create a real-time crash risk assessment model. Others have used Convolutional Neural Netowrk (CNN) to predict the road accident risk map using historical accident data and satellite imagery. The above two works, is not suitable for real-time accident prediction for safety planning. A recent work used the Long ShortTerm Memory (LSTM) model to predict short-term traffic crash risk. However, this work is a simple application of the model on traffic crash data, without addressing the spatial heterogeneity of data.

On the other hand, our project proposes some innovative ideas to solve the problems of spatial heterogeneity in the data we collect. We implement these ideas on an LSTM neural network.

Detection of traffic accident hotspots

Other work on traffic accident analysis includes clustering and hotspot detection. Researchers have proposed a location ratio testing approach to identify road intersections with high traffic accident density. Others have proposed a linear hotspot detection approach to identify paths with significantly higher traffic accident density compared to others. Others have proposed an ontology-based approach, which considers crash severity level for clustering and mapping traffic crash risk.

IV. OVERVIEW

This section presents the data we collected and i the formulation of our problem.

A. Data Sources

We are going to base this on data from Iowa, the study area instead of SHENZEN, given the d encountered when retrieving data from China. Iowa with both rural and urban environments, and various weather conditions (e.g., snowstorm, heavy rain, t Tamerius pointed out that Iowa is ideal for stuc impact of precipitation on traffic crashes because of it weather patterns. All of the data we have collected is as detailed below;

High-Resolution Rainfall Data

We also obtained the stage IV radar rainfall product developed by the NWS. The data contains the amount of hourly precipitation (in millimeters) from the radar at 4 kilometer resolution. There are a total of 8,026 observation tiles, which cover all of Iowa during the study periods.

Motor Vehicle Crash Data

We obtained the traffic accident location, traffic information for 2006 through 2013 from the Iowa Department of Transportation (DOT).

RWIS (Roadway Weather Information System) Data

RWIS is a temperature change monitoring project. The project primarily provides temperature and wind-related characteristics of 86 observing stations located near major highways in Iowa.

Road Networks

We collected three sets of road network data; detailed road speed limits, recent estimates of the annual average daily traffic volume of primary and secondary roads, as well as detailed statistics for each type of vehicle.

Satellite Images

Nous recueillons une image satellite de l'Iowa à partir de Google Earth

Traffic Camera Data

the collection of the number of vehicles circulating in both directions each hour from the cameras of more than 128 HIGHWAY STATIONS .

B. Database

We have the spatial grid S, where each grid si represents a d×d square region.

our goal is to learn a model to predict the total number of crashes in each grid in S during each time slot T . in our case the time slot is every day (24 hours) as the length of t. but we can apply our proposed framework with a different choice of d and t. With time-varying characteristics such as weather, the traffic volume of day t is not available until day t has passed. We therefore only use the characteristics of days up to t -1 to predict accidents on day t. Concretely, we formulate the problem as follows:

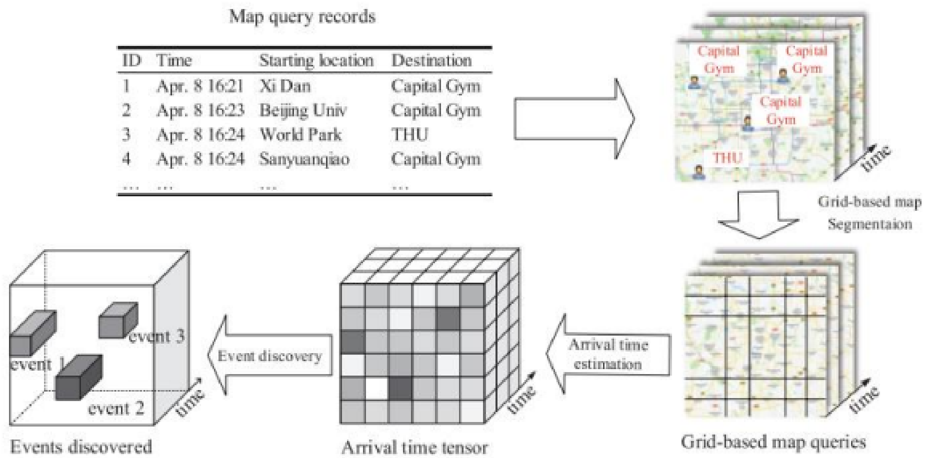

Fig. 1. Q-traffic Dataset

- A spatio-temporal field S × T , where S = s1,s2, ...,sn is a spatial grid, and T = t1,t2, ...,tn is the time length of the study period partitioned into equal-length slots .
- – A 3-D tensor of traffic accident count C, where C(s,t) is the number of accidents in grid s S during day t T
- A list of m feature tensors F = f1, f2, ..., fm , where fk is a three-dimensional tensor recording the corresponding attribute in each grid si during each time slot
- A training data set Dtrain =(C(S,t),F(S,t)) where t Ttrain, and a testing data set Dtest=(C(S,t),F(S,t)) where t T .
- **Find** – A model to predict C(s,t) for every t T
- **Objective** – Minimize prediction erreur
- **Constraints** – The dependency between C and F vary spatially – All accidents occur along the road network - F(S,ti) is not available for predict C(S,ti) , ti Ttest .

V. FEATURE EXTRACTION

To generate the features of our input, we associate the collected datasets with each si ,t combination and aggregate the data to extract a list of features. Dependent Variable Tensor (C): For each grid if on each day t, we count the total number accidents C(Si,t) .We fully matched 375,690 motor vehicle accidents over 8 years.

A. Time-Invariant Features

- **Road Network Mask Features**
  we map the road network (with main and secondary roads) on the grids and create a mask layer. Although the entire study area is gridded, it is evident that traffic accidents can only occur on the road network. To ensure that the prediction results have meaning, we convert the road network into a feature map. Since the road network only changes over the years, this characteristic is time invariant.
- **Road Condition Features RC** in addition to the netmask layer, we calculate the average length of all roads and

the average speed limit of roads in each grid cell and store each of these measurements as a feature. We also include characteristics relevant to road ownership, such as number of intersections, number of lanes, road function, road curve, and annual average daily traffic .

- **Google Earth Satellite Image (G)**
  We get an instant satellite image broken down int color channels (R,G,B) of all of Iowa based latitudes and longitudes of the corners of the imag Google Earth and geo-log the image on Iowa mı each grid cell si for each channel, the feature valu mean value of the overlapping pixels si .

B. *Time-Variant features*

Rainfall Features

we map each radar data tile from the precipitation to the grid cell if in our frame. For each rainfall grid which si , overlaps we find si . All The amount of RA dı tiles si ,t radar is calculated as the average daily total of all these rj tiles on day t. This results in a time-v characteristic.

RWIS Weather feature

unlike traffic accidents, weather characteristics such as temperature are continuously distributed throughout space. Therefore, we find the k nearest RWIS stations to each grid cell si in Euclidean distance. Next, we calculate the average measurement of the three stations at each hour as an estimated hourly measure of si . Finally, we extract the mean hourly measures at si for day t as feature values.

C. *SpatialGraph features*

there are still many factors that could make the accident pattern different in different areas; accidents are concentrated in urban areas than in rural areas, which can be attributed to different population density in different areas. to build a spatial graph between all routes. We get the resulting upper eigenvectors of the Laplacian matrix. These eigenvectors provide additional information about the topological feature of each road with respectively potential space groups in the road. This approach is similar to spectral clustering, it generates the eigenfeatures based on the Laplacian matrix, then conducts k-means clustering based on the new features. We use spectral clustering to visualize the features generated based on the clustering results in Figure 1 (k=10). Finally, we map the entities to each grid cell if . For grid cells with a single road segment, spatialGraph entities are directly affected as those of the corresponding road segment. we select the longest road segment in si and use its spatialGraph functionality for si for grid cells with multiple road segments. In our implementation, we choose k = 10, which results in 10 time-invariant spatialGraph features.

VI. THE HETERO-CONVLSTM APPROACH

The CovnLSTM has nice properties for traffic accident prediction because the LSTM part can capture temporal auto-correlation in the data and the convolution operator can capture local spatial features (e.g. dangerous road intersections) which are important indicators of potential accidents. However, ConvLSTM does not explicitly handle spatial heterogeneity. Although we incorporate the functionality of Spatial Graph.

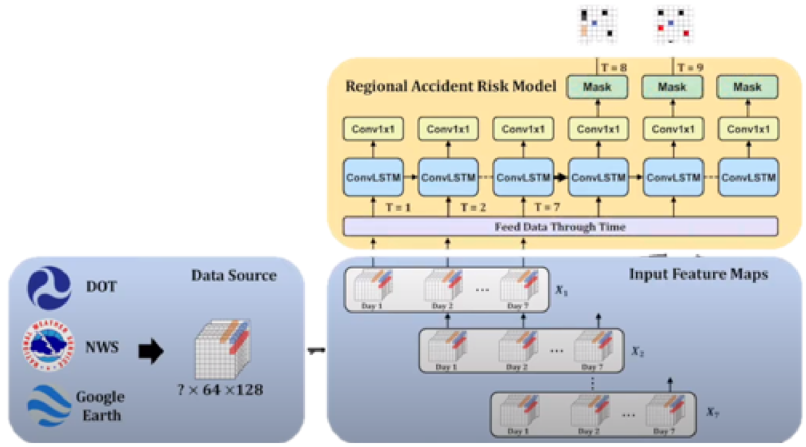

Fig. 2. The structure of the regional ConvLSTM model

All this data is converted into the feature map reconstructing our data grid if in functions of days precisely 7 days because traffic accidents are impacted by human activities, which have a strong weekly pattern. Each training sequence consists of 14 days, where the last 7 days are predicted based on data from the first 7 days (the last 7 days are shifted one day from the first 7 days). All 14-day feature tensors are fed into the ConvLSTM model.

VII. Spatial Ensemble of ConvLSTM Models

The model generates the prediction for day t until day t -1 In order to solve the spatial heterogeneity problem, we build an LSTM model for each of the different regions of the study area and use a sampling method as well as a moving window approach, where the size of the moving window is of 32 × 32. We take subsets of the spatial frame S by moving the window from the upper left corner to the lower right corner, with a pitch of 16 grids on both horizontal and vertical directions. This results in 21 different regions,

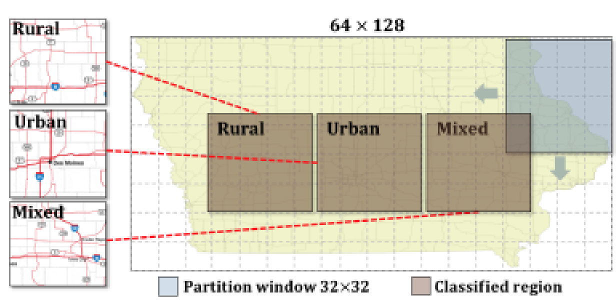

Fig. 3. 5: Map partitioning of spatial ensemble model

VIII. SYSTEM DEPLOYMENT

The project was implemented on the Argon HPC (High Performance Computing) computing system at the University of Iowa using a 256GB RAM compute node with a 2.6GHz 16-core processor. For training deep learning models, they used a GPU node on Argon with Nvidia Tesla P100 accelerator cards with Tensorflow library support.

IX. CONCLUSION

This article was based on a study that was carried out in London with the same objectives, namely the problem of predicting traffic accidents using Deep Learning models on heterogeneous urban data which is an important problem for transport. and public safety. It is also challenging due to the scarcity and spatial heterogeneity of data. This work has shown that Deep Learning techniques such as ConvLSTM are promising solutions for traffic accident prediction. To fully deploy a publicly accessible working system we need to meet the following challenges: Deploy our project to a cloud service/storage for real-time query. Update the model offline periodically to keep predictions accurate. The feasible solution to the first problem is to deploy our models in a cloud service (Amazon AWS Lambda or Microsoft Azure). For the daily forecast, our system will be programmed to run and retrieve the past day's weather, traffic volume, precipitation, and crash data at midnight, turn it into feature maps, and then recycle the model in an online manner.

REFERENCES

[1] https://www.snap4city.org/dashboardSmartCity/management/iframeApp.php?linkUrl=https://www.snap4city.org/drupallinkId=snap4cityPortalLinkpageTitle=www.snap4city.orgfromSubmenu=false
[2] https://dl.acm.org/doi/10.1145/3219819.3219922
[3] https://www.kdd.org/kdd2018/accepted-papers/view/hetero-convlstm-a-deep-learning-approach-to-traffic-accident-prediction-on-
[4] https://dl.acm.org/doi/pdf/10.1145/3219819.3219922
[5] https://github.com/JingqingZ/BaiduTraffic
[6] https://aws.amazon.com/lambda/
[7] http://www.thegazette.com/2013/12/08/snow-blankets-eastern-iowa-causingslick-conditions