Unsupervised LSTMs-based Learning for Anomaly Detection in Highway Traffic Data

Nicola Di Mauro(✉) and Stefano Ferilli

Dipartimento di Informatica – Università di Bari, Bari, Italy
nicola.dimauro@uniba.it

Abstract. Since road traffic is nowadays predominant, improving its safety, security and comfortability may have a significant positive impact on people's lives. This objective requires suitable studies of traffic behavior, to help stakeholders in obtaining non-trivial information, understanding the traffic models and plan suitable actions. While, on one hand, the pervasiveness of georeferencing and mobile technologies allows us to know the position of relevant objects and track their routes, on the other hand the huge amounts of data to be handled, and the intrinsic complexity of road traffic, make this study quite difficult. Deep Neural Networks (NNs) are powerful models that have achieved excellent performance on many tasks. In this paper we propose a sequence-to-sequence (Seq2Seq) autoencoder able to detect anomalous routes and consisting of an encoder Long Short Term Memory (LSTM) mapping the input route to a vector of a fixed length representation, and then a decoder LSTM to decode back the input route. It was applied to the TRAP2017 dataset freely available from the Italian National Police.

Keywords: Traffic Understanding, Autoencoders, Recurrent Neural Networks.

1 Introduction

As a consequence of a much easier and comfortable possibility of traveling nowadays, traffic on roads has become predominant, especially in some Countries, where road traveling is preferred to other options, such as railways or airways. People travel for work, for leisure, or for personal commitments. Actually, a large portion of our lives is spent on roads. As a consequence, roads have become an important component of our lives, and so improving road traffic may have a significant positive impact on our general lives. There are several perspectives on which traffic can be improved: safety (having to do with car accidents), security (related to road crimes), well-being (related to traffic flows, jams and facilities available along the road).

In order to set up appropriate improvements in these fields, a study of traffic behavior is needed. Unfortunately, studying traffic is quite difficult, both because of the lack of publicly available data and because road traffic is much more complex than other means of transportation (e.g., trains or airplanes), due to several reasons:

– possible routes are fixed and quite simple in the other cases, while road vehicles can reach almost any place following any route they like;
– all details of routes, stops, timings, and coordination among vehicles are centrally planned and determined in the other cases, while in road vehicles they depend on a number of factors related to the end-users, and may even dynamically change during the journey;
– the number and variability in type of vehicles traveling on roads is much greater than for other means of transportation.

Hence, automatic techniques that may help stakeholders in carrying out such a study represent a precious resource to obtain the desired goals. For instance, one might want to predict traffic jams, or accidents, in order to place appropriate actions aimed at avoiding such events. Or, one might want to identify abnormal behavior of vehicles, that might be associated to suspect activities. In other cases, one might want to exploit the model to improve the patrolling plans and/or the environment in which traffic takes place, removing criticalities and optimizing the overall behavior.

These techniques may work at various levels: at a lower level, extracting non-trivial information that is relevant to stakeholders, by using data mining approaches, may allow them to react appropriately; at a higher level, obtaining human-readable models that allow stakeholders to have a wide and comprehensive picture of traffic behavior may allow them to identify criticalities and set up more general strategies for improving both the situation and their approach to it. Developing automatic decision support systems that use both the previous levels to carry out some kind of reasoning may help stakeholders in taking their decisions and setting up their plans.

Actual implementation of these techniques is possible today thanks to the pervasiveness of georeferencing and mobile technologies, allowing to both know the static position of relevant objects, and track the routes of moving entities. Not only technology today enables the collection of traffic data at a very fine-grained level. Many institutions and companies actually use this technology: e.g., on urban and extra-urban roads cameras are positioned that can monitor the road situation and car flow; on highways automatic systems can detect the presence of specific cars by reading their plate number—for billing or speed checking purposes; insurance companies provide their customers with GPS trackers; last but not least, everybody uses GPS-enabled mobile phones and navigators that may reveal their position and route.

Several stakeholders have also already started to collect the data sensed by these devices. Outstanding examples are insurance companies and, especially, National Traffic Polices. The Italian National Police, in particular, has recently made available for research purposes a dataset reporting vehicle positions on an Italian highway along one year, in occasion of the 1st Italian Conference on Traffic Police (TRAP2017) [8]. This was very important, because no other datasets with similar features—real data, big data, long detection timespan—are freely available.

In this paper, we propose a sequence-to-sequence (Seq2Seq) [14] deep learning approach based on autoencoder Long Short Term Memory (LSTM) [6] to detect anomalous routes—possibly associated to suspect or otherwise relevant behaviors. A qualitative and quantitative analysis of the experimental results of our approach, applied to the TRAP2017 dataset, shows that the system returns an acceptable number of relevant routes, and that the detected routes may indeed be worth attention by the traffic analysts.

This paper is organized as follows. After introducing some related works in Section 2, subsequent sections describe the reference dataset, the task and the proposed approach. Then, Section 5 discusses experiments and Section 6 concludes the paper.

2 Related Work

The research aimed at extracting information and models from mobility data is a recent and flourishing field in Artificial Intelligence [12]. Works are present focusing on people (persons, groups) or on territory (roads, cities, regions). They aim at providing support and/or recommendation to persons and groups, or helping institutional stakeholders (e.g., police or administration) in planning their activities. Since the approach proposed in this work will be applied to a dataset specifically concerned with highway traffic, and purposely collected to challenge researchers on extracting information that is useful to police activities [9], in the following we will focus on this particular setting and will quickly overview the work carried out so far on this specific topic and dataset, as presented at the First Italian Conference on Traffic Police in 2017.

In the perspective of 'traffic understanding', intended as the general task of obtaining a model of traffic on a given road or in a given region with specific purposes, [4] proposes a process mining approach to learn models of traffic flow that can be used for analysis, supervision, and prediction purposes. [13] suggests to prune the database in order to reduce its size while retaining interesting information, and to apply Formal Concept Analysis to obtain knowledge about the observed behavior patterns in criminal activities.

More oriented to traffic analysis, [1] proposes a scalable and fully automated procedure to effectively import and process traffic data in order to analyze them. After analyzing the TRAP2017 dataset, in [2] the same authors developed a clustering-based tool to model and classify routes and plates, and to identify itineraries possibly related to criminal intents. Their tool also allows some characterization of the clusters, useful for police officers to better understand them.

Other works focus on prediction. Although based on a different dataset, [10] presents a tensor completion based method for highway traffic prediction that can take into account multi-mode features (such as daily and weekly periodicity, spatial information, temporal variations, etc.). Trying to overcome the drawbacks of traditional models and algorithms that require a predefined and static length of past data, and do not take into account dynamic time lags and temporal autocorrelation, [5] explores the use of LSTMs obtaining higher accuracy

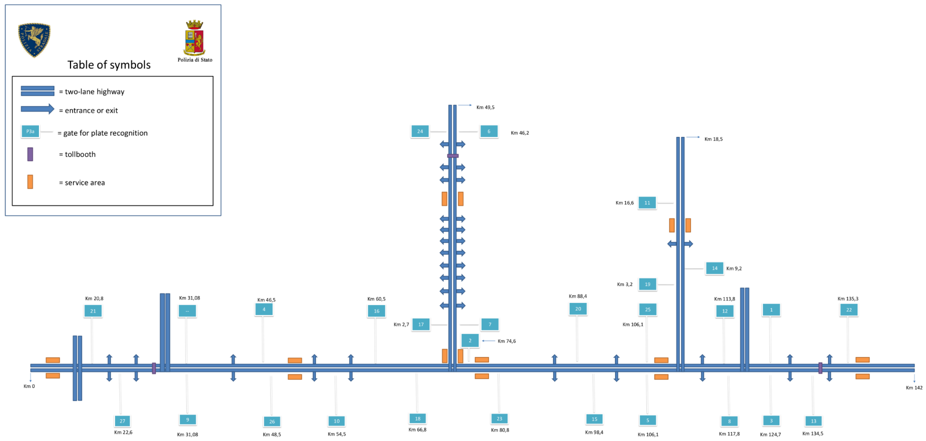

Fig. 1. Map of the highway considered by TRAP2017.

and good generalization. [11] proposed a stacked autoencoder model for traffic flow prediction, which considers the spatial and temporal correlations inherently. Differently from [11, 5], here we propose a Seq2Seq autoencoder based on LSTM in order to study anomalous routes.

3 Dataset Description and Task Definition

The TRAP2017 dataset includes a log of transits of vehicles on a limited portion of Italian highways. The data were collected in 365 days (year 2016) by the Italian National Police using automatic Number Plate Reading Systems placed on 27 'gates' spread along the road. Each of the 155,586,309 entries in the log reports: plate number (anonymized), gate id, lane id, timestamp, and plate nationality. An (anonimyzed) map of the considered portion of highways is shown in Figure 1, reporting gates (along with their position expressed in kilometers), entrances/exits, tollbooths, and service areas.

Among all possible tasks to be carried out on these data, we focused on spotting routes that can be considered anomalous for some reason, and thus may require further analysis. Since the dataset is organized by day of the year, we defined a route as the temporal sequence of gates passed by a given plate number in a day. As a consequence, we may have routes of different length. Since the dataset is not annotated, we decided to go for an unsupervised approach, and so from a technical viewpoint we set up an outlier detection task. This means that anomalous routes are not interpreted, and our definition for what is anomalous is simply something unexpected.

We aimed at proposing a general technique that may be applied to many situations. For this reason, we decided to use as few information as possible. Specifically, we decided to completely ignore the map, since it may not be available in some cases, and to only focus on the sequence of gates, ignoring all the other fields in the dataset—time, lane and nationality. In particular, we ignore time, which is expected to be a very relevant feature, both as regards how long it takes to go from a gate to the next one—too long or too short times may indicate anomalous behavior—and as regards the day, month or season—different days and periods of the year are typically associated to different traffic behaviors.

4 Proposed Approach

The Seq2Seq setting has received significant research attention. It first uses an encoder to encode a source sequence, and then applies a decoder to the encoded sequence in order to decode it into a target sequence. The goal is to estimate the conditional probability of generating the target sequence given the encoding of the source sequence.

4.1 Autoencoders

An autoencoder neural network, first introduced as a dimensionality reduction method, is an unsupervised learning algorithm that applies backpropagation, setting the target values $\mathbf{y}$ to be equal to the inputs $\mathbf{x}$. Internally, it has a hidden layer $\mathbf{h}$ that describes a code used to represent the input. In particular, it transforms an input vector $\mathbf{x}$ into a latent representation $\mathbf{h}$—the encoded representation—as follows:

$$\mathbf{h} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b}) \tag{1}$$

where $\mathbf{W}$ and $\mathbf{b}$ correspond to the weights and bias in the encoder, and σ is the sigmoid function. The latent representation $\mathbf{h}$ is then mapped back into the reconstructed feature space as follows:

$$\mathbf{y} = \sigma(\mathbf{W}'\mathbf{h} + \mathbf{b}') \tag{2}$$

where $\mathbf{W}'$ and $\mathbf{b}'$ correspond to the weights and bias in the decoder. The autoencoder is trained by minimizing the reconstruction error $||\mathbf{y} - \mathbf{x}||$. Many autoencoders can be connected to build a stacked autoencoder, which can be used to learn multiple levels of non-linear features.

4.2 Recurrent Neural Network

Recurrrent Neural Networks (RNNs) are a generalization of feedforward NNs to sequences. Given an input sequence $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$, a standard RNN iteratively computes an output sequence $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T)$ using the following equations:

$$h_t = f(Ux_t + Wh_{t-1}) \tag{3}$$

$$y_t = Vh_t, \tag{4}$$

where f is usually a nonlinearity such as tanh or Rectified Linear Unit (ReLU) activation function, being $\mathbf{h}$ the internal state of the RNN.

A RNNs Seq2Seq based approach consists in using a RNN to map an input sequence to a fixed-sized vector and then map such a vector to the output sequence with another RNN [3]. However, learning with this approach might be difficult due to the resulting long term dependencies, a problem solved by adopting Long Short-Term Memory (LSTM)—a special kind of RNN. LSTM overcome the vanishing gradients problem by replacing an ordinary neuron by a complex architecture called the LSTM unit or block.

4.3 The model

In our proposed autoencoder LSTM model, the encoder is a LSTM NN that reads each symbol of an input sequence $\mathbf{x}$ sequentially. After reading the end of the sequence, the hidden state of the LSTM NN is a summary $\mathbf{c}$ of the whole input sequence $\mathbf{x}$. The decoder is another LSTM NN whose initial hidden state is set to the representation $\mathbf{c}$ of $\mathbf{x}$, and trained to generate the output sequence $\mathbf{y}$ [14, 3]. Specifically, the input to the encoder is a sequence of integers (denoting the gates), each encoded as one-hot vectors with length 28 (the number of gates). Since here the aim is to correctly reconstruct the input, the decoder aims at reconstructing the same sequence of integers as the input. An anomaly is reported when the input sequence has not been correctly reconstructed by the autoencoder. In such a case, the system can also report an estimate of the difference between the actual sequence and the predicted one, which can be straightforwardly interpreted as the degree of anomaly of the actual sequence itself. In this way the autoencoder can be used to distinguish between normal and anomalous routes. The proposed model was implemented in Python using the Keras Deep Learning library[1]. The weights of the model, minimizing a categorical cross-entropy, were learned using the adam optimizer [7] for a number of epochs provided as input.

5 Experiments

We ran several experiments, with routes of different length $n \in [7, 13]$. Experiment associated to length n aimed at detecting anomalous routes of length n. This range was selected as representative of the more frequent route lengths, and allowed us to check the behavior of our approach on different inputs. The number of epochs for training was set to 5, as a trade-off between expected effectiveness and efficiency. After splitting the dataset into 70% training and 30% test sets, we ran our experiments from two different perspectives:

analytic aimed at extracting relevant routes in historical data;
predictive aimed at determining whether a new route, not included in historical data, is worth attention.

In the former perspective, autoencoding-based anomaly detection was applied to the training routes; in the latter perspective, it was applied to routes in the test set.

[1] `https://keras.io/`

Table 1. Statistics on the dataset and experimental results.

	Length	7	8	9	10	11	12	13
	# Routes	2124515	1893074	733406	539718	308938	284695	260271
Training	# Routes	1487160	1325151	513384	377802	216256	199286	182189
	% Accuracy	96.52	97.51	98.73	96.67	92.61	93.99	94.84
	# Anomalous	51753	32996	6520	12581	15981	11977	9401
Test	# Routes	637355	567923	220022	161916	92682	85409	78082
	% Accuracy	96.48	97.49	98.57	96.49	92.37	93.65	94.35
	# Anomalous	22435	14255	3146	5683	7072	5423	4412

5.1 Quantitative Analysis of Results

Statistics on the dataset and quantitative experimental results, divided by route lengths, are reported in Table 1. Specifically, the number of routes (overall, used for training, and used for testing) is reported in rows labeled '# Routes'. For training and test outcomes, corresponding to the 'analytic' and 'predictive' settings respectively, both the percentage of correct reconstuctions (% Accuracy) and the absolute number of anomalous routes proposed (# Anomalous, corresponding to the complement of Accuracy).

We first note that the percentages of accuracy on training and test sets are quite similar, indicating that we may expect to have the same performance on unseen routes as we have on known ones. This is important, because the 'predictive' setting is expected to be adopted continuously by the police officers, and so it must ensure high-quality results on which basing their decisions. Then, we note that the percentages of anomalous routes identified for the various lengths are very low, and that they are lower for lengths involving more routes. This is important in order to keep to a minimum the burden on police officers that are supposed to check these routes. The largest number of anomalous routes is about 70000, corresponding to length 7. This amounts to about 5850 a month, i.e., about 195 per day on average, which may be considered an acceptable rate.

5.2 Qualitative Analysis of Results

Let us now enter into more detail about the quality of our results, by showing some sample anomalous routes identified by the system. Specifically, in Table 2 we reported three representative anomalous routes for each possible length: the most anomalous one—having the highest reconstruction error estimate, the most frequent one—having the largest number of occurrences in the set of anomalous routes, and one which is apparently strange. Due to space considerations, we cannot discuss all of them in depth. However, the reader may use the map in Figure 1 to have an idea of the routes, and of how strange they may be considered.

Consider, for instance, the following routes:

- route (4, 26, 10, 4, 26, 4, 27, 9, 26, 4, 26, 4, 26). It contains many loops between gates 4 and 26, with occasional passages from other gates. Since a

Routes length	Anomalous routes
7	(15, 15, 3, 27, 5, 27, 1) (9, 27, 4, 16, 2, 20, 25) (9, 26, 10, 23, 2, 16, 4)
8	(27, 4, 16, 2, 20, 25, 1, 22) (27, 4, 16, 2, 20, 12, 1, 22) (21, 27, 4, 16, 2, 20, 25, 12)
9	(1, 1, 22, 25, 1, 22, 25, 12, 1) (4, 16, 2, 20, 25, 15, 23, 10, 26) (20, 20, 16, 16, 4, 4, 10, 18, 23)
10	(23, 2, 2, 18, 23, 2, 23, 2, 23, 2) (15, 23, 18, 10, 9, 21, 4, 16, 2, 20) (4, 16, 2, 20, 14, 5, 23, 18, 26, 9)
11	(12, 3, 5, 1, 5, 25, 12, 1, 25, 12, 1) (27, 26, 10, 23, 15, 5, 25, 20, 2, 16, 4) (22, 1, 12, 20, 2, 23, 15, 5, 8, 3, 13)
12	(6, 24, 4, 26, 6, 24, 4, 26, 4, 16, 17, 24) (1, 12, 25, 22, 1, 12, 25, 1, 12, 25, 1, 25) (5, 15, 23, 18, 10, 9, 21, 4, 16, 2, 20, 25)
13	(4, 26, 10, 4, 26, 4, 27, 9, 26, 4, 26, 4, 26) (10, 18, 23, 5, 8, 3, 13, 22, 1, 20, 2, 16, 4) (13, 3, 8, 5, 23, 18, 10, 16, 2, 25, 12, 1, 22)

Table 2. Three representative anomalous routes for each route length in [7,13].

service station is present exactly between those gates, there is a chance that this car is repeatedly stopping in this station, possibly to look for possible people to be robbed;

– another quite strange route is (20, 20, 16, 16, 4, 4, 10, 18, 23), where there are several pairs of occurrences of the same gate, and these gates are not close to each other. This raises the question about why a car should alternate in-highway and out-highway routes aimed at passing twice from the same place, and for very distant places. This is especially interesting since sometimes there is no apparent reason (e.g., a service station) for passing from those places;
– route (23, 2, 2, 18, 23, 2, 23, 2, 23, 2) basically consists of many passages from two gates—23, 2—without ever passing from intermediate gates that are present on the highway.

It is also interesting to note that some gates very frequently appear in the strange routes, while some other appear very seldom or do not appear at all. This is evident also in the tiny fraction of anomalous routes selected in Table 2. E.g., gates 2 and 20 are present in about 2/3 of the selected routes, and gate 2 in particular occurs 16 times in them. Also some combinations are more frequent and might deserve attention. E.g., pair (2,20) occurs in nearly half the selected routes; pair (26,4) occurs several times in two routes, in spite of these two gates being placed on opposite sides (and direction) on the highway, indicating that the vehicle performed several loops in a day on that portion of the road.

6 Conclusions and Future Work

Improving safety, security and comfort of road traffic may have a dramatic positive impact on people's lives, because most of our time is spent on roads. To effectively pursue this objective, stakeholders must rely on suitable studies of traffic behavior that provide them with non-trivial information, so that they may understand the traffic models and plan suitable actions. While, on one hand, the pervasiveness of georeferencing and mobile technologies allows us to know the position of relevant objects and track their routes, on the other hand the huge amounts of data to be handled, and the intrinsic complexity of road traffic compared to other means of transportation, make this study quite difficult. This calls for automatic techniques based on Artificial Intelligence approaches.

This paper proposed an approach to traffic mining aimed at identifying potentially anomalous behavior that is worth attention, and applied it to the TRAP2017 dataset, concerning a year's data about highway traffic, freely available from the Italian National Police. A quantitative analysis of the results showed that the number of anomalous routes identified is consistent with our expectations, and a qualitative analysis of the results revealed that the routes selected by our approach are indeed peculiar, and deserve further insight. Future work will proceed in several directions. A cooperation with police operators that may evaluate the outcomes and provide expert insight is planned. Also, a study of the relationships between the procedure's parameters and the quantity an quality of anomalous routes returned will be carried out. Finally, we plan to design further modules that, in cooperation with the proposed one, will provide police officers with further information on the available traffic data. If the good performance of the approach will be confirmed, we will build a system that will support police officers in their activities.

Acknowledgments

Work partially funded by the Italian PON 2007-2013 project PON02_00563_3489339 'Puglia@Service'.

References

1. Bernaschi, M., Celestini, A., Guarino, S., Lombardi, F., Mastrostefano, E.: Traffic data: Exploratory data analysis with apache accumulo. In: Leuzzi and Ferilli [8], pp. 129–144

2. Bernaschi, M., Celestini, A., Guarino, S., Lombardi, F., Mastrostefano, E.: Unsupervised classification of routes and plates from the *trap2017* dataset. In: Leuzzi and Ferilli [8], pp. 97–114
3. Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder–decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1724–1734 (2014)
4. Ferilli, S., Redavid, D.: A process mining approach to the identification of normal and suspect traffic behavior. In: Leuzzi and Ferilli [8], pp. 37–56
5. Fumarola, F., Lanotte, P.F.: Exploiting recurrent neural networks for gate traffic prediction. In: Leuzzi and Ferilli [8], pp. 145–153
6. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation 9(8), 1735–1780 (1997), `https://doi.org/10.1162/neco.1997.9.8.1735`
7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR abs/1412.6980 (2014), `http://arxiv.org/abs/1412.6980`
8. Leuzzi, F., Ferilli, S. (eds.): 1st Italian Conference for the Traffic Police (TRAP-2017). Advances in Intelligent Systems and Computing, Springer (2017)
9. Leuzzi, F., Del Signore, E., Ferranti, R.: Towards a pervasive and predictive traffic police. In: Leuzzi and Ferilli [8], pp. 19–36
10. Liao, J., Zhao, X., Tang, J., Zhang, C., He, M.: Efficient and accurate traffic flow prediction via fast dynamic tensor completion. In: Leuzzi and Ferilli [8], pp. 69–82
11. Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.: Traffic flow prediction with big data: A deep learning approach. IEEE Trans. Intelligent Transportation Systems 16(2), 865–873 (2015), `https://doi.org/10.1109/TITS.2014.2345663`
12. Nanni, M.: Advancements in mobility data analysis. In: Leuzzi and Ferilli [8], pp. 11–16
13. Rodriguez-Jimenez, J.: Detecting criminal behaviour patterns in Spain and Italy using Formal Concept Analysis. In: Leuzzi and Ferilli [8], pp. 57–68
14. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: NIPS14. pp. 3104–3112 (2014)