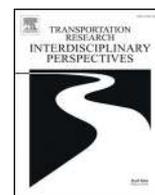




Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: <https://www.journals.elsevier.com/transportation-research-interdisciplinary-perspectives>



A framework for end-to-end deep learning-based anomaly detection in transportation networks

Neema Davis^{*}, Gaurav Raina, Krishna Jagannathan

Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai 600 036, India



ARTICLE INFO

Article history:

Received 23 October 2019

Received in revised form 12 March 2020

Accepted 11 April 2020

Available online 8 May 2020

Keywords:

End-to-end anomaly detection

LSTM

Extreme value theory

ABSTRACT

We develop an end-to-end deep learning-based anomaly detection model for temporal data in transportation networks. The proposed EVT-LSTM model is derived from the popular LSTM (Long Short-Term Memory) network and adopts an objective function that is based on fundamental results from EVT (Extreme Value Theory). We compare the EVT-LSTM model with some established statistical, machine learning, and hybrid deep learning baselines. Experiments on seven diverse real-world data sets demonstrate the superior anomaly detection performance of our proposed model over the other models considered in the comparison study.

© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The increasing availability of large-scale traffic data sets provides an opportunity for innovation in Intelligent Transportation Systems. The avenues for exploration are numerous, ranging from uncovering traffic patterns (Lippi et al., 2010), city dynamics (Zheng et al., 2011), driving directions (Yuan et al., 2010), discovering demand hot spots (Chang et al., 2010), finding vacant taxis around a city (Phithakkitnukoon et al., 2010), predicting taxi demand (Davis et al., 2019b), taxi operation patterns (Li et al., 2011), to detecting anomalies (Chen et al., 2013), among others.

The various verticals of Intelligent Transportation Systems have received adequate research attention in the past. However, the recent emergence of deep learning techniques and their applicability in transportation systems has resulted in a heightened interest in this area (Wang et al., 2019). Consequently, traditional machine learning models in many applications are now being replaced by deep learning techniques, which is reshaping the landscape of intelligent transport networks. An application domain that has benefited significantly from the evolution of deep learning-based technologies is anomaly detection (Chalapathy, 2019). Anomaly detection aims to find patterns that are not normally expected from the data. Typical observations from traffic data demonstrate strong spatio-temporal patterns, showing periodicity and correlations between adjacent locations. These patterns may vary depending on the time of the day, day of the week, season, or location. Occasional deviations from these patterns can be termed as abnormal events. While various

short-term forecasting models can learn about periodic patterns in the data (Davis et al., 2018), they usually are unable to capture these anomalous events accurately. However, it is necessary to detect these unusual events as well as they often indicate useful and critical information that can yield instructive insights. For example, abnormal traffic event detection can be utilized to mitigate congestion, plan driving routes, and reduce taxi demand-supply imbalance.

Based on the nature of data, anomalous event detection can find uses in various applications. Within the transportation domain, anomaly detection has been applied to abnormal trajectory detection (Chen et al., 2013), finding atypical regions (Kong et al., 2018), obstacle detection (Dairi et al., 2018), congestion analysis (Markou et al., 2017), and irregularities in taxi passenger demand (Wittmann et al., 2018), among others. One also finds extensive use in a wide range of applications such as fraud detection for credit cards, insurance, or health care, intrusion detection for cybersecurity, fault detection in safety-critical systems, and military surveillance for enemy activities (Chandola et al., 2009).

1.1. Related literature

Traditionally, anomaly detection has been performed using parametric and non-parametric statistical models, data clustering, rule-based systems, mixture models, and SVMs (Support Vector Machines), among others. For extensive surveys, the interested reader can refer to (Chandola et al., 2009) and (Hodge and Austin, 2004). These traditional models often fail to fully capture complex structures in the data. Additionally, as the volume of data increases, traditional methods may experience difficulties in finding outliers at large scale. Thus, the performance of the aforementioned

^{*} Corresponding author.

E-mail addresses: ee14d212@ee.iitm.ac.in, (N. Davis), gaurav@ee.iitm.ac.in, (G. Raina), krishnaj@ee.iitm.ac.in. (K. Jagannathan).

algorithms, in detecting outliers, might be sub-optimal for emerging real-world use cases.

In recent years, deep learning-based anomaly detection algorithms have become increasingly popular, with applications in a diverse set of tasks (Chalapathy, 2019). Unsupervised anomaly detection using deep learning has mainly been hybrid in nature. First, the deep neural network learns the complex patterns of the data. Then, the hidden layer representations from this trained network are used as input to traditional anomaly detection algorithms. There are two popular categories of deep learning-based hybrid anomaly detection. The first category consists of methods that analyze the reconstruction errors in an auto-encoder trained over the normal data. A deficiency in the reconstruction of a test point indicates abnormality (Malhotra et al., 2016). The second class of methods utilizes either an auto-encoder trained over the normal class to generate a low-dimensional embedding, or a neural network to generate predictions. To identify anomalies, one applies classical methods over the embedding or predictions, such as a parametric distribution assumption (Malhotra et al., 2016), an OC-SVM (One Class-SVM) (Oza and Patel, 2018), etc.

While the currently popular hybrid deep learning-based anomaly detection techniques have proven to be effective in multiple tasks, these neural networks are not customized for anomaly detection. Since the hybrid models extract features using a neural network and feed it to a separate anomaly detection method, they fail to influence the representational learning in the hidden layers. A more advanced variant of this approach combines the encoding and detection steps using an appropriate objective function, which is used to train a single neural model that performs both procedures (Ruff et al., 2018). In another related research (Golan and El-Yaniv, 2018), the authors use geometrical transformations to perform end-to-end deep learning-based anomaly detection using CNNs (Convolutional Neural Networks). In (Chalapathy et al., 2018), an OC-SVM objective is implemented in a feed-forward neural network for deep anomaly detection.

The primary focus of the aforementioned literature is on anomaly detection in the context of image data sets. The anomaly detection techniques tailored for images need not necessarily perform well with time-sequences. Therefore, in this study, we aim to develop an end-to-end anomaly detection using LSTM (Long Short-Term Memory) network (Gers et al., 1999), which is a neural network designed for sequential data. By gathering insights from EVT (Extreme Value Theory) (Siffer et al., 2017), we design an end-to-end LSTM-based anomaly detection model. To the best of our knowledge, an LSTM-based end-to-end deep anomaly detection model for transportation data has not been explored in the literature. Further, our objective function and network weight update rules are based on results from EVT. So far, Extreme Value Theory has not been employed in training a neural network model for performing anomaly detection. These features set our research apart from existing literature.¹

1.2. Our contributions

We propose an end-to-end deep anomaly detection algorithm, and compare the model against several baseline models: (i) parametric GARCH (Generalized Auto Regressive Conditional Heteroskedasticity) model, (ii) non-parametric OC-SVM model, and (iii) hybrid LSTM anomaly detection models based on different detection rules. The detection rules used in the hybrid deep anomaly detection models are based on Gaussian distribution assumptions, Tukey's method, and EVT. An evaluation across seven diverse data sets shows that our proposed EVT-LSTM model outperforms the traditional statistical, machine learning, and hybrid deep learning baseline models. Through this study, we highlight the need for a customized neural network model in a deep learning-based anomaly detection setting.

The rest of the paper is organized as follows. In Section 2, we explain the traditional baseline models considered for anomaly detection in this study. The hybrid deep anomaly detection model, along with the three detection

strategies, is explained in Section 3. It is followed by Section 4, where we introduce our proposed EVT-LSTM model. The experimental settings are provided in Section 5, and the results are outlined in Section 6. We conclude our work in Section 7.

2. Traditional anomaly detection

In this section, we provide brief descriptions of two traditional anomaly detection models considered as baselines in our comparison study.

2.1. GARCH model

Parametric statistical models (Fox, 1972) represent one of the early works on outlier detection in time-series. Several models were subsequently proposed in the literature for parametric anomaly detection, including ARMA (Auto Regressive Moving Average), ARIMA (Auto Regressive Integrated Moving Average), and EWMA (Exponentially Weighted Moving Average), to list a few (Chandola et al., 2009). We assume that the normal data instances are located at the high probability regions of a stochastic model compared to the anomalies that have a low probability. A common practice followed here is to either assume a distribution for the anomalies (Eskin, 2000) or fit a regression model to the data (Chen et al., 2005).

A regression-based anomaly detection technique involves two steps: (a) fit a regression model to the data, (b) the residuals, *i.e.*, the part not explained by the regression model, are used to determine the anomaly scores. A popular choice for regression-based anomaly detection is the GARCH model (Engle, 2001), which is typically applied to financial time-series. A GARCH process is often preferred over other regression models such as ARMA because it imposes a specific structure on the conditional variance of the process. The variance is not assumed to be a constant, making the series non-stationary in nature and rendering them suitable for real-world scenarios. Essentially, the GARCH process models the error variance of the time-series as an ARMA process. The AR part models the variance of the residuals and the MA part models the variance of the process. The time-series ϵ_t at each instance t is given by:

$$\epsilon_t = \sigma_t w_t, \quad (1)$$

where, w_t is discrete white noise with zero mean and unit variance, and σ_t^2 is given by:

$$\sigma_t^2 = \delta_0 + \sum_{i=1}^r \delta_i \sigma_{t-i}^2 + \sum_{i=1}^s \gamma_i \epsilon_{t-i}^2, \quad (2)$$

where, δ_i and γ_i are the parameters of the model. In other words, ϵ_t is a Generalized Auto Regressive Conditional Heteroskedastic model of order r and s , denoted by GARCH(r, s).

Parametric methods allow the model to be evaluated very rapidly for new instances and are suitable for large data sets; the model grows only with model complexity and not the data size. However, they limit their applicability by enforcing a predetermined distribution to the data. These approaches are accurate only if the data fits the chosen distribution model. The non-parametric approach described below can overcome this disadvantage associated with parametric models.

2.2. OC-SVM model

Non-parametric methods such as SVMs (Schölkopf et al., 2002) apply local kernel models rather than a single global distribution model to the data. Their popularity stems from the ability to combine speed and low complexity growth of parametric methods with the model flexibility associated with non-parametric methods. Kernel-based methods estimate the density distribution of the input space and identify outliers as lying in regions of low density.

Typically, the SVM model is given a set of training examples labeled as belonging to one of two classes. The model tries to divide the training

¹ A part of this work has been presented as a conference paper (Davis et al., 2019a, 2019b).

sample points into two categories by creating a boundary while penalizing those training samples that fall on the wrong side of the boundary. The SVM model can then make predictions by assigning points to either side of the boundary. For anomaly detection applications, the training examples are often limited. As a result, SVMs are more popularly applied in a one-class setting here, where the SVM model is trained on data that has only one class, that is the *normal* class. This is particularly useful in anomaly detection because by inferring the properties of the normal class, the examples that deviate from the normal class can be identified. The SVM model needs a kernel function that can map the original non-linear observations into a higher-dimensional space in which they are separable. Commonly used kernel functions are Linear, Sigmoid, Gaussian, and RBF (Radial Basis Function) (Schölkopf et al., 2002, Chapter 2). During the testing phase, if a test instance falls within the learned region, it is declared as normal, else it is deemed as anomalous.

The SVM model requires a kernel function, which has to be carefully tuned for obtaining good classification accuracy. Further, the anomaly detection is supervised in nature; it requires prior knowledge of the labels. On the other hand, the recently developed anomaly detection models based on neural networks can perform unsupervised anomaly detection, and hence, have seen widespread use over SVM models for anomaly detection lately.

3. Hybrid deep anomaly detection

Neural networks can perform unsupervised modeling and learn complex time-sequences, which makes them suitable candidates for anomaly detection in large real-world data sets. When we feed non-anomalous data to the network, the model learns the normal behavior of the system. Later, when the network encounters an instance that deviate significantly from the rest of the data, the model classifies it as an anomalous event. This classification is performed on the basis of the prediction errors generated from the model. A high prediction error associated with any sample indicates the presence of a potential anomaly. For performing such prediction error-based anomaly detection, we need *detection rules*. Detection rules are usually traditional statistical or machine learning-based algorithms applied on top of the prediction errors to generate anomaly scores. Anomalous data points are detected based on these anomaly scores and a suitable threshold. In other words, prediction error-based deep anomaly detection is hybrid in nature. Popular detection rules involve thresholding the prediction errors (Wang et al., 2011), making distributional assumptions (Malhotra et al., 2016), or applying machine learning techniques such as an SVM model to the errors (Ergen et al., 2017). We now briefly describe the prediction model and detection rules considered in our study.

3.1. Prediction model

The LSTM model is a popular neural network architecture widely used for learning temporal sequences (Gers et al., 1999). This motivates the choice of an LSTM network in our anomaly detection setting, as we aim to learn from temporal data and use the predictions to perform detection. At each instant, l_b number of values are fed into the model to create l_a number of predictions. The parameters l_b and l_a are known as look-back and look-ahead times respectively. Further, to avoid over-fitting, techniques such as dropout and early stopping are applied.

We divide each data set into training, validation, and test sets. The model learns from the training data, and validates its performance on the hold-out validation data. We assume that the training set is free of anomalies. This is a reasonable assumption as instances of normal behavior are often available in abundance, but instances of anomalous behavior may be rare in real-world use cases. Validation and test sets are comprised of both normal and anomalous samples. First, we feed the training data, *i.e.*, the data without anomalies, to the neural network. This enables the model to learn the normal behavior of the data. Once training is completed, the model utilizes the validation set to derive an appropriate anomaly threshold for that data. Then, the model detects anomalies from a test set

by applying this threshold to the prediction errors obtained. In this work, we define prediction error at time t as the absolute difference between the input and the corresponding network output at t .

For developing anomaly thresholds on prediction errors, we consider three detection rules: (i) the Gaussian-based technique that assumes a Normal distribution on the prediction errors, (ii) the Tukey's method that does not make any assumptions on the distribution, and (iii) the EVT-based rule that assumes a tail distribution, but makes no assumptions about the parent distribution. Those instances for which the corresponding prediction error values lie beyond the chosen threshold are classified as anomalies. The detection rules are explained below.

3.2. Gaussian-based detection (Malhotra et al., 2016)

One of the earliest and popular works in prediction-based anomaly detection setting (Malhotra et al., 2016) assumes that the prediction errors from the training set follow a Gaussian distribution. The prediction errors obtained from the LSTM model are fit to a Gaussian distribution. The mean, μ , and variance, σ^2 , of the distribution are computed using MLE (Maximum Likelihood Estimation) (Myung, 2003). The Log PDs (Probability Densities) of errors are calculated based on the parameters estimated, and they act as anomaly scores. Lower the value of Log PD, higher is the likelihood that the observation is an anomaly. A threshold τ_g is determined on the Log PD values based on a validation set containing both normal and anomalous data. The threshold is chosen such that it incurs as few false alerts as possible while detecting all the abnormal instances from the set of observations. A separate test set is used to evaluate the chosen threshold.

3.3. Tukey's method based detection (Wang et al., 2011)

Tukey's method uses percentiles to set anomaly thresholds without making any distributional assumptions. That is, no quantitative measures, such as the moments of the function, are required to classify the data. In Tukey's method, an instance is marked as an outlier if it lies outside the threshold $\tau_t = Q_3 + 3 \times (Q_3 - Q_1)$, where Q_1 is the lower quartile or the 25th percentile, and Q_3 is the upper quartile or the 75th percentile. The metric $Q_3 - Q_1$ is known as the interquartile distance. After obtaining the prediction errors from the training, validation, and test sets, the errors are concatenated to calculate the lower quartiles and interquartile distances. Any value lying outside τ_t is identified as a potential outlier.

3.4. EVT-based detection (Siffer et al., 2017)

For a random variable X , the CDF (Cumulative Distribution Function) is defined as $F(x) = P(X \leq x)$. Similarly, $\bar{F}(x) = P(X > x)$ denotes the tail distribution. The probability $P(X > x)$ tends to zero for the extreme events in the system. A key result from EVT (De Haan and Ferreira, 2007) suggests that the distribution of the extreme deviations in any system is not highly sensitive to the parent distribution. In other words, we can accurately compute the probabilities of extreme values without estimating the underlying distribution. Under a weak condition, the extreme events have the same kind of distribution, regardless of the parent distributions, known as the EVD (Extreme Value Distribution):

$$G_{(\sigma, \gamma)} : y \rightarrow \exp \left(- \left(1 + \gamma \frac{y}{\sigma} \right)^{-\frac{1}{\gamma}} \right), \quad \gamma \in \mathbb{R}, \quad 1 + \gamma \frac{y}{\sigma} > 0, \quad (3)$$

where, σ is the scale parameter and γ is the extreme value index of the distribution. Based on the value γ takes, the tail distribution can be Fréchet ($\gamma > 0$), Gumbel ($\gamma = 0$), or Weibull ($\gamma < 0$). Once we associate an EVD with the tail of an unknown distribution, the probability of the extreme events can be readily computed. Recently, results from EVT have been applied to the problem of anomaly detection in uni-variate data streams (Siffer et al., 2017). In that study, the authors follow the POTs (Peaks-

Over-Thresholds) approach to design a threshold. Rather than fitting an EVD to the extreme values of X , the POTs approach fits a GPD to the excesses $X-T$, where T is some initial threshold. Often, T is chosen as the 98% quantile. To compute parameter estimates for GPD, we follow the procedure outlined by (Grimshaw, 1993). Once the parameters are obtained, the threshold τ_e can be computed as:

$$\tau_e = T + \frac{\hat{\sigma}}{\hat{\gamma}} \left(\left(\frac{qn}{N_t} \right)^{-\hat{\gamma}} - 1 \right), \quad (4)$$

where, $\hat{\sigma}$ and $\hat{\gamma}$ are the estimated parameters of the GPD, q is the desired probability or the risk, n is the number of observations, and N_t is the number of peaks, i.e., the number of X_i s.t. $X_i > T$. The probability $P(X > \tau_e)$ is calculated for the test set, and those data instances for which $P(X > \tau_e) < q$ are classified as plausible anomalies. The authors in (Siffer et al., 2017) recommend choosing a value for q within $[10^{-3}, 10^{-5}]$, which we follow in our study. More details of this algorithm can be found in (Siffer et al., 2017).

4. End-to-end deep anomaly detection

In Section 1.1, we highlighted the need for developing end-to-end deep learning-based anomaly detection models, especially for time-series data. An end-to-end deep anomaly detection technique involves modifying the objective function of a deep learning model such as an LSTM or a CNN. Modifications are introduced so that the models that were formerly learning patterns for forecasting will now learn to detect deviations from the normal behavior. Instead of first predicting using a neural network and then feeding the predictions to a separate post-processing technique, the outputs of an end-to-end deep anomaly detection model can be directly interpreted as anomaly scores. In (Ruff et al., 2018), the authors combine a CNN with an SVDD (Support Vector Deep Description) objective. The SVDD is a technique similar to the OC-SVM, where a hyper-sphere is used to separate the data instead of a hyper-plane.

Let $\phi(\cdot; \mathcal{W}) : \mathcal{X} \rightarrow \mathcal{Y}$ be a neural network with L layers and a set of weights $\mathcal{W} = \{\mathbf{W}^1, \dots, \mathbf{W}^L\}$. This network maps data from an input space $\mathcal{X} \subseteq \mathbb{R}^p$ to an output space $\mathcal{Y} \subseteq \mathbb{R}^q$. That is, $\phi(\mathbf{x}; \mathcal{W}) \in \mathcal{Y}$ is the network representation of $\mathbf{x} \in \mathcal{X}$ given by the network ϕ with parameters \mathcal{W} . The One-Class Deep SVDD objective given in (Ruff et al., 2018), for a CNN model with input $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, is as follows:

$$\min_{\mathcal{W}} \frac{1}{N} \sum_{i=1}^N \|\phi(\mathbf{x}_i; \mathcal{W}) - \mathbf{c}\|^2 + \frac{\lambda}{2} \sum_{l=1}^L \|\mathbf{W}^l\|_F^2. \quad (5)$$

The first term in the quadratic loss objective function penalizes the distance between every network representation $\phi(\mathbf{x}_i; \mathcal{W})$ and the center of the hyper-sphere \mathbf{c} . The second term penalizes the network weights by employing a network weight decay regularizer with hyper-parameter $\lambda > 0$, where $\|\cdot\|_F$ denotes the Frobenius norm. In (Ruff et al., 2018), the \mathbf{c} was fixed as the mean of the network predictions that results from performing an initial forward pass on the training data samples. The experiments were conducted for MNIST and CIFAR-10 image data sets.

In order to develop a similar model for time-sequences, we implement the aforementioned objective function in an LSTM model. Interestingly, we find that while this quadratic loss objective function works satisfactorily for anomaly detection in images, it does not fare well for temporal data. When adopted in the LSTM network, we notice that Eq. (5) minimizes the distance between the predictions and their initial mean by reducing the magnitude of the predictions, resulting in a large fraction of false positives. This behavior suggests that an objective function that directly minimizes the network predictions might not be a sensible choice for anomaly detection in temporal data. We recall that the success of hybrid deep learning-based anomaly detection algorithms was mainly attributed to an efficient threshold based on the prediction errors. Therefore, it is natural to explore an objective function that minimizes the prediction errors and not the actual predictions.

Further, in our recent work (Davis et al., 2019a), after comparing different detection strategies for hybrid deep anomaly detection, we noticed the potential of a strategy based on extreme values. We found that an EVT-based detection rule performed better than other popular detection techniques. The superior performance of an EVT-based strategy in a deep learning setting encouraged us to integrate EVT into the objective function of the LSTM model, leading to an end-to-end deep anomaly detection model.

4.1. EVT-LSTM model

In our study, the inputs $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ in $\mathcal{X} \subseteq \mathbb{R}^p$ are mapped to the set $\{y_1, \dots, y_N\}$ in $\mathcal{Y} \subseteq \mathbb{R}$. Our EVT-LSTM model is based on the objective function given as follows:

$$\min_{\mathcal{W}} \frac{1}{N} \sum_{i=1}^N \|\mathbb{E}(\phi(\mathbf{x}_i; \mathcal{W})) - \tau_e\|^2 + \frac{\lambda}{2} \sum_{l=1}^L \|\mathbf{W}^l\|_F^2. \quad (6)$$

Here, instead of minimizing the distance between the network representations and the mean obtained after an initial forward pass as in Eq. (5), we minimize the Euclidean distance between every absolute prediction error $\mathbb{E}(\phi(\mathbf{x}_i; \mathcal{W}))$ and a threshold τ_e . The threshold τ_e is obtained from Eq. (4), and is updated periodically during the training phase. This form of optimization is called an alternating minimization approach and has been used with similar objective functions in related literature (Ruff et al., 2018; Chalapathy et al., 2018). The objective functions in these related literature minimized a function of the predictions obtained from image data sets. On the other hand, our objective function in Eq. (6) optimizes a function of the prediction errors. Our proposed algorithm is given in Algorithm 1.

Algorithm 1. The training process of the proposed EVT-LSTM model. The threshold τ_e is updated every $k = 20$ epochs.

Input: Set of examples (\mathbf{x}_n, y_n) , $n \in \{1, \dots, N\}$
Output: Set of decision scores s_n , $n \in \{1, \dots, N\}$
Initialization: Threshold $\tau_e \leftarrow 0$
while convergence criteria unmet **do**
 Update weights of the network using Eqn. (6)
 for once in every k epochs **do**
 Calculate prediction errors, $\mathbb{E}(\hat{y}_n) = |\hat{y}_n - y_n|$
 $T \leftarrow \text{InitThreshold}(\mathbb{E}(\hat{y}_n))$
 Excesses $\leftarrow \{\mathbb{E}(\hat{y}_n) - T \mid \mathbb{E}(\hat{y}_n) > T\}$
 Fit a GPD to excesses by using MLE and find $\hat{\gamma}$, $\hat{\sigma}$
 Update τ_e using Eqn. (4)
 end
end
Compute decision score $s_n = |\hat{y}_n - y_n| - \tau_e$ for each \mathbf{x}_n
if $s_n \geq 0$ **then**
 | \mathbf{x}_n is anomalous
else
 | \mathbf{x}_n is non-anomalous
end

The threshold τ_e is initialized to zero at the beginning of the experiment. During the training phase, the LSTM model tries to optimize the objective function given in Eq. (6). The prediction errors on the training set are calculated every k epochs. The 98% empirical quantile of the errors is chosen to set an initial threshold T in $\text{InitThreshold}(\mathbb{E}(\phi(\mathbf{x}; \mathcal{W})))$. The excesses occurring above T are fit to a GPD using MLE, and the parameters $\hat{\gamma}$ and $\hat{\sigma}$ are estimated. Then, using Eq. (4), we calculate the new value for the threshold τ_e . The objective function in Eq. (6) is updated with this recent value of threshold obtained. The next k epochs use the modified objective function to train the model, after which the threshold τ_e is again calculated and updated. The training stops when either the convergence is achieved, or the maximum

number of epochs is reached. Finally, on a test set, the decision scores are calculated and used for classifying instances as anomalous or non-anomalous.

5. Experimental settings

In this section, we discuss the data sets considered, evaluation metrics used, and the procedure for choosing parameters for each anomaly detection model.

5.1. Description of data sets

Seven diverse real-world data sets are considered in our comparison study. We employ three road traffic-based data sets, two taxi demand data sets, and two data sets from miscellaneous application domains. The travel time, vehicle occupancy, and traffic speed data sets considered are real-time data obtained from a traffic sensor near the Twin Cities Metro area and collected by the Minnesota Department of Transportation. These traffic data sets are available at the Numenta Anomaly Benchmark GitHub repository.² The NYC (New York City) taxi demand data set is publicly available at (Taxi and Limousine Commission, 2016) and contains the trip details of government-run street hailing taxis. The proprietary Bengaluru taxi demand data set is acquired from a leading private Indian transportation company dealing with app-based taxi-hailing services. The ECG (electrocardiogram) data is available in (Keogh et al., 2005) and has annotations from a cardiologist to indicate the unusual heartbeat patterns. Bitcoin historical prices are obtained from coindesk³ package, R.

Brief descriptions of the data sets used are given below.

1. Vehicular Travel Time²: The travel time data is obtained from a traffic sensor and has 2500 readings from July 10, 2015, to September 17, 2015, with eight marked anomalies. The sensor outputs are obtained from a road link, in intervals of 10 min.
2. Vehicular Speed²: The data set contains the average speed of all vehicles passing through the traffic detector, obtained in 5 min intervals. A total of 1128 readings for the period September 8, 2015 - September 17, 2015, is available. There are three marked unusual sub-sequences in the data set.
3. Vehicle Occupancy²: There are a total of 2382 readings indicating the percentage of the time, during a 30-s period, that the detector sensed a vehicle. The data is available for a period of 17 days, from September 1, 2015, to September 17, 2015, and has two marked anomalies. A reading is obtained once in every 5 min, from a traffic sensor for a road link.
4. NYC (New York City) Taxi Demand (Taxi and Limousine Commission, 2016): The publicly available NYC data set contains the pick-up locations and time stamps of street hailing yellow taxi services from the period of January 1, 2016, to February 29, 2016. The data is aggregated over 15 min time intervals in 1 km² grids. This spatio-temporal aggregation results in sequences of length 5760 from more than 700 grids. We pick three time-sequences (S1, S2, and S3) with clearly apparent anomalies from data aggregated.
5. Bengaluru Taxi Demand: This data set has GPS traces of passengers booking a taxi by logging into the service provider's mobile application. Similar to the NYC data set, this data is also available for January and February 2016. We aggregate the data over 15 min periods in 1 km² grids and pick three sequences with clearly visible anomalies. Similar to NYC data set, we have time-sequences, each of length 5760, from around 740 grids.
6. ECG (Electrocardiogram) (Keogh et al., 2005): There are a total of 21,600 readings, with three unusual sub-sequences labeled as anomalies. The data set has a repeating pattern, with some variability in the period length.
7. Bitcoin Prices³: Historical bitcoin prices are available for the period from January 1, 2017, to May 27, 2019. The fraction of anomalies in this data

Table 1

Appropriate ARIMA(p, d, q)-GARCH(r, s) models obtained for each data set, by varying p, q in the range [0, 5], d in [0, 1], and r, s in [1, 2]. The anomaly thresholds are obtained from a hold-out validation set, so that as few false positives are incurred.

Data Sets	Model	Threshold
Vehicular Travel Time	ARIMA(1, 0, 3)-GARCH(1, 1)	0.016
Vehicular Speed	ARIMA(0, 1, 4)-GARCH(1, 1)	0.036
Vehicle Occupancy	ARIMA(0, 1, 1)-GARCH(1, 1)	0.433
NYC Taxi Demand	S1 ARIMA(0, 1, 3)-GARCH(1, 1)	0.009
	S2 ARIMA(3, 0, 4)-GARCH(1, 1)	0.047
	S3 ARIMA(2, 1, 2)-GARCH(2, 2)	0.051
Bengaluru Taxi Demand	S1 ARIMA(1, 0, 3)-GARCH(1, 2)	0.064
	S2 ARIMA(3, 1, 3)-GARCH(1, 1)	0.003
	S3 ARIMA(1, 0, 1)-GARCH(1, 2)	0.060
Electrocardiogram	ARIMA(4, 1, 2)-GARCH(1, 1)	10 ⁻⁶
Bitcoin Prices	ARIMA(2, 1, 2)-GARCH(1, 1)	0.025

set of 877 readings is observed to be 0.06%, most of them occurring around the beginning of the year 2018.

5.2. Evaluation metrics

We consider three evaluation metrics for comparing our models: (i) Precision, P , (ii) Recall, R , and (iii) F1-score, $F1$, which is the harmonic mean of Precision and Recall. Min-max normalization is performed on every data set before modeling and evaluation.

1. Precision, P :

$$P = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}, \quad (7)$$

2. Recall, R :

$$R = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}, \quad (8)$$

3. F1-score, $F1$:

$$F = 2 \cdot \frac{P \times R}{P + R}. \quad (9)$$

True positives are the anomalous instances that have been correctly classified as anomalies by the model. Similarly, true negatives are the instances correctly identified as non-anomalous data. False positives are the non-anomalies incorrectly classified as anomalous, and false negatives are the incorrectly identified anomalies. Since F1-score summarizes both Precision and Recall, we consider the model with the highest F1-score as the superior anomaly detection technique. For each data set, we manually select the validation and test sets so that they contain both anomalous and non-anomalous instances while preserving the seasonality in the sequences. The training-validation-test split is approximately 60–10–30.

5.3. Parameter selection

In order to perform efficient anomaly detection, it is necessary to set appropriate hyper-parameters and anomaly thresholds for each model. The suitable set of parameters and thresholds vary with the use case considered. Below, we briefly discuss the procedures through which the parameters are shortlisted for each anomaly detection model.

² <https://github.com/numenta/NAB/tree/master/data>

³ <https://cran.r-project.org/package=coindesk>

Table 2

The shortlisted OC-SVM models for the data sets considered. We consider Linear, Sigmoid, Polynomial, and RBF kernels, and vary α between [0.0001, 0.1].

Data Sets	Kernel Setting
Vehicular Travel Time	RBF(0.0001)
Vehicular Speed	Poly(0.0001)
Vehicle Occupancy	RBF(0.0001)
NYC Taxi Demand	RBF(0.0001)
	S1
	S2
	S3
Bengaluru Taxi Demand	RBF(0.0001)
	S1
	S2
	S3
Electrocardiogram	Linear
Bitcoin Prices	Sigmoid(0.0001)

5.3.1. GARCH model

For every data set, time-sequences are generated based on the training data. For Bengaluru and NYC taxi demand data sets, the temporal aggregation is performed at sampling periods of 15 min. Then, by varying the p , q , and d parameters of an ARIMA(p , d , q) process between [1, 5], appropriate models are chosen for every time-sequence. The residuals obtained from fitting the ARIMA processes are then modeled as suitable GARCH(r , s) processes. We find that suitable values for parameters r and s often lie in the range [1, 2]. Once appropriate models are developed, anomaly scores are obtained based on the deviation of the GARCH predictions from the actual values. An anomaly threshold is set based on the validation set and examined on a test set. The parameters of the fitted ARIMA-GARCH models, along with the anomaly thresholds are given in Table 1.

5.3.2. OC-SVM model

Appropriate kernel functions are crucial for satisfactory anomaly detection performance of SVMs, and the choices vary with the data sets considered. In our study, we consider Linear, RBF, Polynomial, and Sigmoid kernels. Another important parameter is the kernel coefficient α for the RBF, Polynomial, and Sigmoid kernels. After varying α in the range [0.0001, 0.1], a value of 0.0001 is found to suit most of the data sets considered. For every use case, we ran multiple SVM models on the training data, with different parameters chosen from the range of values considered. Then, suitable choices are made by observing the classification accuracy on a hold-out validation set. Finally, the best OC-SVM model obtained is used to detect anomalies on a test set. The shortlisted OC-SVM models are given in Table 2.

5.3.3. Hybrid LSTM models

Before training any neural network model, it is essential to set appropriate hyper-parameter values. These parameters cannot be inferred while

Table 3

The LSTM architectures for the data sets considered. For each data set, hyper-parameters are chosen after running the TPE (Tree-structured Parzen Estimator) Bayesian Optimization algorithm.

Data Sets	LSTM Architecture
Vehicular Travel Time	1 Recurrent layer: {20}, Dropout: 0.2, 1 Dense layer: {1}, Learning rate: 0.01
Vehicular Speed	1 Recurrent layer: {60}, Dropout: 0.19, 1 Dense layer: {1}, Learning rate: 0.0001
Vehicle Occupancy	1 Recurrent layer: {50}, Dropout: 0.23, 1 Dense layer: {1}, Learning rate: 0.0001
NYC Taxi Demand	2 Recurrent layers: {50, 20}, Dropout: 0.4, 1 Dense layer: {24}, Learning rate: 0.0001
Bengaluru Taxi Demand	2 Recurrent layers: {20, 10}, Dropout: 0.25, 1 Dense layer: {24}, Learning rate: 0.0001
Electrocardiogram	2 Recurrent layers: {60, 30}, Dropout: 0.1, 1 Dense layer: {5}, Learning rate: 0.05
Bitcoin Prices	1 Recurrent layer: {10}, Dropout: 0.1, 1 Dense layer: {1}, Learning rate: 0.0001

Table 4

The chosen false positive regulator values for the LSTM-based hybrid anomaly detection models. While the thresholds for both Gaussian and Tukey's method based models vary significantly with each data set considered, the probability values for EVT-based detection is found to remain within $[10^{-3}, 10^{-5}]$.

Data Sets	Hybrid LSTM Models		
	Gaussian (τ_g)	Tukey (τ_t)	EVT (q)
Vehicular Travel Time	-20	572.9	10^{-4}
Vehicular Speed	-18	24.4	10^{-3}
Vehicle Occupancy	-23	12.9	10^{-5}
NYC Taxi Demand	S1 -19	12.1	10^{-5}
	S2 -17	12.8	10^{-5}
	S3 -15	10.5	10^{-5}
Bengaluru Taxi Demand	S1 -25	33.5	10^{-4}
	S2 -18	27.1	10^{-4}
	S3 -25	14.0	10^{-4}
Electrocardiogram	-23	0.1	10^{-4}
Bitcoin Prices	-17	12,961.8	10^{-3}

training the model as they correspond to the model selection task and influence the speed of the learning process. Hyper-parameters pertaining to the model selection task include topology and size of the network. Similarly, mini-batch size, drop out, and learning rate are some of the hyper-parameters that affect the speed and quality of the learning process. Since the selection of the suitable hyper-parameters is not a trivial task, we employ a Bayesian Optimization technique known as the TPE (Tree-structured Parzen Estimator) algorithm (Bergstra et al., 2011) for this exercise. We consider a fully connected dense layer with Relu activation as the output layer. The objective to be minimized is the Mean Squared Error, for which we use the Adam optimizer (Kingma and Ba, 2014). All the LSTM-based models are trained for 100 epochs, with a mini-batch size of 64.

The hyper-parameters shortlisted for each data set are given in Table 3. For the ECG data, we follow the architecture given in (Singh, 2017), where the authors have shortlisted suitable parameters for this data set. The limited availability of readings for data sets such as traffic speed, travel time, vehicle occupancy, and bitcoin prices suggest small look-back and look-ahead periods. Since we have over 10 million points for the two taxi demand data sets, we can have a larger look-back time. For scenarios with large l_b , we find that the LSTM learns better representations of the data, aiding the anomaly detection process.

A key parameter that can influence the accuracy of the detection algorithms is the false positive regulator. For the Gaussian or Tukey's method based detection rule, this parameter is the corresponding threshold τ_g or τ_t . For EVT-based techniques, the desired probability q is the false positive regulator. We select a threshold τ_g for the Gaussian-based hybrid anomaly detection such that the threshold maximizes the F1-score on the validation set. The threshold τ_t for the Tukey's method is directly obtained from the entire set of prediction errors, based on a simple quantile calculation. For both hybrid and end-to-end EVT-LSTM deep learning models, we follow similar procedures to set the parameters for EVT rule. Prediction errors from the training and validation sets are concatenated to form an initialization data stream. The desired probability q for the EVT-based rule is chosen based on this data stream. We use the same sequence to choose the initial threshold T , which is typically set to 98% quantile. The false positive regulator q is chosen such that the EVT algorithm detects all the anomalous instances from the data stream. The chosen values for the false positive regulators of the hybrid LSTM-based techniques are given in Table 4.

5.3.4. EVT-LSTM model

The hyper-parameters and false positive regulators chosen for hybrid LSTM models are used for the EVT-LSTM model as well. We follow the guidelines in (Ruff et al., 2018) while setting the hyper-parameter λ for the network weight regularizer. The threshold is updated every $k = 20$ epochs. The values chosen for hybrid deep learning models seem to suit end-to-end deep learning models, for most of the scenarios considered. An exception is the Bengaluru taxi demand data set, where the suitable

Table 5

P-values obtained on statistical testing. Null hypothesis is rejected if the p-values are found to lie below 0.001. In all the data sets considered, the null hypothesis that the tail distribution follows a GPD is accepted.

Data Sets		P-values
Vehicular Travel Time		0.005
Vehicular Speed		0.005
Vehicle Occupancy		0.370
NYC Taxi Demand	S1	0.805
	S2	0.056
	S3	0.147
Bengaluru Taxi Demand	S1	0.570
	S2	0.180
	S3	0.006
Electrocardiogram		0.002
Bitcoin Prices		0.051

value for q turns out to be 10^{-5} . Nevertheless, the best choices for the probability q remain within $[10^{-3}, 10^{-5}]$.

6. Results

In this section, we analyze whether the tails of the prediction error distributions follow a GPD, and present results from the numerical tests performed.

6.1. Statistical tests

We conduct a statistical test known as the A-D (Anderson-Darling) test (Stephens, 1974) to check whether the tail distribution follows a GPD. The A-D test can be used to assess whether a sample of the data comes from a particular probability distribution. This test makes use of the specific distribution while calculating the critical values. The test statistic A^2 calculates the distance between the hypothesized distribution and the empirical CDF of the data. The null hypothesis states that the data follow a specific distribution, which is GPD in our case. Based on the test static and the p-values obtained, the null hypothesis can (cannot) be rejected. The A-D test is a modification of the KS (Kolmogorov-Smirnov) test (Massey Jr, 1951), and gives more weight to the tails than does the KS test. We conduct the A-D test on the excesses $X-T$, i.e., on the prediction errors lying above the empirical threshold T . The p-values obtained from this statistical test are given in Table 5. We reject the null hypothesis for each data set if the corresponding p-value lies below 0.001. From the table, we find that the prediction error distribution tail appears to follow GPD, for all the data sets considered. This finding supports our proposal to employ an EVT-based detection rule.

Table 6

The anomaly detection performance of various models considered in the study, across diverse data sets, based on F1 -score. The proposed end-to-end EVT-LSTM deep anomaly detection model is observed to perform better compared to the statistical, machine learning and hybrid deep learning techniques considered.

Data Sets	Anomaly Detection Models						
		GARCH	OC-SVM	LSTM Tukey (Hybrid)	LSTM Gaussian (Hybrid)	LSTM EVT (Hybrid)	EVT-LSTM (End-to-End)
Vehicular Travel Time		0.01	0.04	0.07	0.21	0.36	0.36
Vehicular Speed		0.18	0.56	0.79	0.74	0.79	0.79
Vehicle Occupancy		1.0	0.33	0.5	1.0	1.0	1.0
NYC Taxi Demand	S1	0.002	0.03	0.25	1.0	1.0	1.0
	S2	0.005	0.16	0.14	0.33	1.0	1.0
	S3	0.007	0.6	0.66	0.86	0.86	0.86
Bengaluru Taxi Demand	S1	0.03	0.29	0.47	0.57	1.0	1.0
	S2	0.002	0.12	0.08	0.5	0.5	0.66
	S3	0.04	0.44	0.26	0.54	0.62	0.72
Electrocardiogram		0.1	0.22	0.49	0.32	0.37	0.28
Bitcoin Prices		0.52	0.31	0.19	0.83	0.83	0.84

6.2. Numerical results

Table 6 contains the anomaly detection performance of various models across different data sets, based on the F1-score metric. We draw the following inferences from the table:

- The poor performance of the parametric GARCH models suggest that assuming a particular distribution on the prediction errors can critically affect anomaly detection accuracy.
- Deep learning-based anomaly detection algorithms exhibit superior detection accuracy over statistical and machine learning-based algorithms across seven diverse data sets.
- Out of the two classes of deep learning-based anomaly detection models considered, an end-to-end detection algorithm outperforms hybrid detection models on a broad variety of data sets.

When the parametric GARCH model is employed for anomaly detection, we observe that the model has a sufficiently high Recall, but very low Precision. The threshold chosen based on the validation set classifies a large number of non-anomalies as anomalous on the test set. Thus, the overall anomaly detection performance is affected by the presence of several false positives, resulting in a low F1-score value. Exceptions to this behavior are observed with vehicle occupancy data set and to an extent, with the bitcoin prices data. The magnitude of the anomalies is much higher than that of the non-anomalies in these data sets, which appears to be the reason behind this exception.

The OC-SVM model achieves a higher detection accuracy compared to statistical GARCH model but does not fare well compared to the deep learning variants. They also showcase high Recall and poor Precision values. On the other hand, a single value of kernel coefficient α (0.0001) proved to be a satisfactory fit for all the data sets considered.

On comparing hybrid and end-to-end deep anomaly detection models, we see that the proposed end-to-end EVT-LSTM model shows superior detection accuracy. The anomaly detection requires no post-processing tools, and the performance is always at least as good as that of the hybrid models considered, for the majority of data sets considered. This observation suggests that a deep learning model customized for anomaly detection can provide better accuracy results than running traditional algorithms on a deep learning model developed for forecasting. The only exception is observed in the ECG data set, which can be attributed to the anomaly labeling scheme followed. The labeling scheme employed in this data set marks an entire period of the ECG signal as anomalous in case any point in that period is an anomaly. In other words, we deal with *collective* anomalies in this data set. The fraction of anomalies is, hence, higher in the ECG data set compared to other data sets that have point anomalies. Thus, the anomalies cover a broad spectrum above the upper quartile of prediction errors for the ECG data. Since the Tukey's method thresholds the raw prediction errors based on the upper quartile, it results in good anomaly detection for

the ECG data set. This finding suggests that a simple threshold based on the magnitude of prediction errors might be sufficient when the fraction of anomalies in the data set is relatively high. Generally, Tukey's method detects most of the anomalies but results in a large number of false positives, similar to GARCH and OC-SVM models. This behavior is not desirable in an anomaly detection setting.

An important observation is made regarding the variability in false positive regulator values of various methods. Recalling the results from Table 4, we find high variability in the false positive regulator values of Gaussian and Tukey detection rules. The choices for thresholds τ_g and τ_e vary significantly with the data set considered. While τ_g varies between $[-15, -25]$, τ_e is found to take values between $[0.11, 12,961.8]$. The dependence of the anomaly thresholds on the time-sequence considered limits the applicability of such detection rules. On the other hand, the only free parameter for EVT-based detection, the probability q , does not appear to have a significant dependence on the data set. This false positive regulator is found to stay within the range $[10^{-3}, 10^{-5}]$. A false positive parameter with low dependency on the data sets is highly preferred in real-world settings, thereby strengthening the case of a detection algorithm based on EVT.

The accuracy improvement obtained by our proposed model on data sets originating from different verticals of transportation suggests the generalization capability of the end-to-end deep learning-based EVT-LSTM model. The model can be used in conjunction with a broad range of data sets. Further, it requires no separate post-processing techniques, which is a clear advantage over the popular hybrid deep learning-based anomaly detection models. Our model is also unsupervised in nature and requires no anomaly labels.

In summary, considering data sets from various verticals of transportation networks, we found that an end-to-end deep learning-based anomaly detection algorithm holds great potential in detecting abnormal traffic instances. Our proposed EVT-LSTM model accurately detected anomalous traffic speed, vehicle occupancy, travel time, and taxi demand instances, in addition to data sets from medical and financial domains.

7. Conclusions

We conclude with a summary of our contributions, and some avenues for further research.

7.1. Contributions

In this study, we explored anomaly detection techniques for various transportation based data sets such as traffic speed, travel time, vehicular occupancy, and taxi demand, among others. Detection of anomalies can aid Intelligent Transportation Systems by providing recommendations for better road network and traffic management. For example, the sudden drops in traffic speed can be indicators of road accidents. The detection of such events can facilitate the timely intervention of officials to deal with possible emergencies. An unusual increase in travel time may be related to traffic congestion. These events may occur even on very short timescales, and can be utilized by the drivers to follow less congested routes. Anomaly detection also finds applications in ride-hailing taxi services. Unexpected spikes in the demand for taxis may be used by taxi service providers to re-route additional drivers to meet the sudden rise in demand. Motivated by the role of anomalous event detection in such scenarios, we developed an end-to-end deep learning-based algorithm that performs unsupervised and near real-time anomaly detection for transportation networks.

The key contributions of this paper are outlined below.

- We proposed a novel end-to-end deep anomaly detection algorithm for temporal data that incorporates concepts from EVT (Extreme Value Theory) into the objective function of an LSTM (Long Short-Term Memory) deep learning model.
- The proposed EVT-LSTM model does not require additional post-processing techniques. The output network representations from our model can be directly utilized for anomaly detection, which is a clear

advantage over the currently popular hybrid deep learning-based detection models that require separate post-processing tools.

- We observed the superior anomaly detection performance of the EVT-LSTM model across seven diverse data sets, by comparing against established statistical, machine learning, and hybrid deep learning baseline models. The proposed model was able to detect true positives faithfully while incurring as few false positives as possible.
- We highlighted the need for a customized neural network model in a deep learning-based anomaly detection setting.

7.2. Avenues for future research

There are numerous avenues that merit further attention. For additional validation of the proposed algorithm, new data sets can be introduced. While our algorithm employs an objective function based on EVT, it would be useful to explore other objective functions, to enhance the detection accuracy. Further, it is necessary to identify and quantify the factors that cause anomalies. The work in (Huang et al., 2018) employs a Markov machine to model traffic speed and volume data and use mutual information to find the potential causes for anomalies. Correlation matrices (Hojati et al., 2016) and decision trees (Zhang and Chen, 2019) have been utilized to measure the impact of accidents and weather effects on travel time. While several deep learning models have used features to improve the prediction accuracy of a variable (Koesdwiady et al., 2016), there are limited investigations on the influences of various input features on anomalous events in deep learning-based models.

References

- Bergstra, J.S., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems*, pp. 2546–2554.
- Chalapathy, R., Chawla, S., 2019. Deep learning for anomaly detection: A survey, arXiv preprint arXiv:1901.03407.
- Chalapathy, R., Menon, A.K., Chawla, S., 2018. Anomaly detection using one-class neural networks, arXiv preprint arXiv:1802.06360.
- Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection: a survey. *ACM Comput. Surv.* 41, 15–73.
- Chang, H.-w., Tai, Y.-c., Hsu, J.Y.-j., 2010. Context-aware taxi demand hotspots prediction. *Int. J. Bus. Intell. Data Mining* 5, 3–18.
- Chen, D., Shao, X., Hu, B., Su, Q., 2005. Simultaneous wavelength selection and outlier detection in multivariate regression of near-infrared spectra. *Anal. Sci.* 21, 161–166.
- Chen, C., Zhang, D., Castro, P.S., Li, N., Sun, L., Li, S., Wang, Z., 2013. iBOAT: isolation-based online anomalous trajectory detection. *IEEE Trans. Intell. Transp. Syst.* 14, 806–818.
- Dairi, A., Harrou, F., Sun, Y., Senouci, M., 2018. Obstacle detection for intelligent transportation systems using deep stacked autoencoder and k -nearest neighbor scheme. *IEEE Sensors J.* 18, 5122–5132.
- Davis, N., Raina, G., Jagannathan, K., 2018. Taxi demand forecasting: a hedge-based tessellation strategy for improved accuracy. *IEEE Trans. Intell. Transp. Syst.* 19, 3686–3697.
- Davis, N., Raina, G., Jagannathan, K., 2019a. LSTM-based anomaly detection: Detection rules from extreme value theory. *Proceedings of the EPIA Conference on Artificial Intelligence*. Springer, pp. 572–583.
- Davis, N., Raina, G., Jagannathan, K., 2019b. Grids versus graphs: Partitioning space for improved taxi demand-supply forecasts, arXiv preprint arXiv:1902.06515.
- De Haan, L., Ferreira, A., 2007. *Extreme Value Theory: An Introduction*. Springer Science & Business Media.
- Engle, R., 2001. GARCH 101: the use of ARCH/GARCH models in applied econometrics. *J. Econ. Perspect.* 15, 157–168.
- Ergen, T., Mirza, A.H., Kozat, S.S., 2017. Unsupervised and semi-supervised anomaly detection with LSTM neural networks, arXiv preprint arXiv:1710.09207.
- E. Eskin, Anomaly detection over noisy data using learned probability distributions, in: *Proceedings of the International Conference on Machine Learning*, Morgan Kaufmann Publishers Inc., 2000, pp. 255–262.
- Fox, A.J., 1972. Outliers in time series. *J. R. Stat. Soc. Ser. B Methodol.* 34, 350–363.
- Gers, F.A., Schmidhuber, J., Cummins, F., 1999. Learning to forget: Continual prediction with LSTM. *Proceedings of the International Conference on Artificial Neural Networks, IET*, pp. 850–855.
- Golan, I., El-Yaniv, R., 2018. Deep anomaly detection using geometric transformations. *Advances in Neural Information Processing Systems*, pp. 9758–9769.
- Grimshaw, S.D., 1993. Computing maximum likelihood estimates for the generalized pareto distribution. *Technometrics* 35, 185–191.
- Hodge, V., Austin, J., 2004. A survey of outlier detection methodologies. *Artif. Intell. Rev.* 22, 85–126.
- Hojati, A.T., Ferreira, L., Washington, S., Charles, P., Shobeirinejad, A., 2016. Modelling the impact of traffic incidents on travel time reliability. *Transportation Research Part C: Emerging Technologies* 70, 86–97.
- Huang, T., Liu, C., Sharma, A., Sarkar, S., 2018. Traffic system anomaly detection using spatiotemporal pattern networks. *Int. J. Prognostics Health Manag.* 9, 1–12.

- Keogh, E., Lin, J., Fu, A., 2005. Hot sax: Efficiently finding the most unusual time series subsequence. *Proceedings of the International Conference on Data Mining*, IEEE, pp. 1–8.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.
- Koesdwiady, A., Soua, R., Karray, F., 2016. fv. *IEEE Trans. Veh. Technol.* 65, 9508–9517.
- Kong, X., Song, X., Xia, F., Guo, H., Wang, J., Tolba, A., 2018. LoTAD: long-term traffic anomaly detection based on crowdsourced bus trajectory data. *World Wide Web* 21, 825–847.
- Li, B., Zhang, D., Sun, L., Chen, C., Li, S., Qi, G., Yang, Q., 2011. Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. *Proceedings of the International Conference on Pervasive Computing and Communications Workshops*. IEEE, pp. 63–68.
- Lippi, M., Bertini, M., Frasconi, P., 2010. Collective traffic forecasting. *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 259–273.
- Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., Shroff, G., 2016. LSTM-based encoder-decoder for multi-sensor anomaly detection, arXiv preprint arXiv:1607.00148.
- Markou, I., Rodrigues, F., Pereira, F.C., 2017. Use of taxi-trip data in analysis of demand patterns for detection and explanation of anomalies. *Transp. Res. Rec.* 2643, 129–138.
- Massey Jr., F.J., 1951. The kolmogorov-smirnov test for goodness of fit. *J. Am. Stat. Assoc.* 46, 68–78.
- Myung, I.J., 2003. Tutorial on maximum likelihood estimation. *J. Math. Psychol.* 47, 90–100.
- N. Y. C. Taxi & Limousine Commission, 2016. TLC trip record data. <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page> accessed 2019-10-01.
- Oza, P., Patel, V.M., 2018. One-class convolutional neural network. *IEEE Signal Proc. Lett.* 26, 277–281.
- Phithakkitmukoon, S., Veloso, M., Bento, C., Biderman, A., Ratti, C., 2010. Taxi-aware map: Identifying and predicting vacant taxis in the city. *Proceedings of the International Joint Conference on Ambient Intelligence*. Springer, pp. 86–95.
- Ruff, L., Görnitz, N., Deecke, L., Siddiqui, S.A., Vandermeulen, R., Binder, A., Müller, E., Kloft, M., 2018. Deep one-class classification. *Proceedings of the International Conference on Machine Learning*, pp. 4390–4399.
- Schölkopf, B., Smola, A.J., Bach, F., 2002. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT press.
- Siffer, A., Fouque, P.-A., Termier, A., Largouet, C., 2017. Anomaly detection in streams with extreme value theory. *Proceedings of the International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 1067–1075.
- Singh, A., 2017. Anomaly Detection for Temporal Data Using Long Short-Term Memory, Master's Thesis, KTH Royal Institute of Technology.
- Stephens, M.A., 1974. EDF statistics for goodness of fit and some comparisons. *J. Am. Stat. Assoc.* 69, 730–737.
- Wang, C., Viswanathan, K., Choudur, L., Talwar, V., Satterfield, W., Schwan, K., 2011. Statistical techniques for online anomaly detection in data centers. *Proceedings of the International Symposium on Integrated Network Management and Workshops*, IEEE, pp. 385–392.
- Wang, Y., Zhang, D., Liu, Y., Dai, B., Lee, L.H., 2019. Enhancing transportation systems via deep learning: a survey. *Transportation Research Part C: Emerging Technologies* 19, 144–163.
- Wittmann, M., Kollok, M., Lienkamp, M., 2018. Event-driven anomalies in spatiotemporal taxi passenger demand. *Proceedings of the International Conference on Intelligent Transportation Systems*. IEEE, pp. 979–984.
- J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, Y. Huang, T-drive: Driving directions based on taxi trajectories, in: *Proceedings of the SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, 2010, pp. 99–108.
- Zhang, X., Chen, M., 2019. Quantifying the impact of weather events on travel time and reliability. *J. Adv. Transp.* 2019, 1–9.
- Y. Zheng, Y. Liu, J. Yuan, X. Xie, Urban computing with taxicabs, in: *Proceedings of the International Conference on Ubiquitous Computing*, ACM, 2011, pp. 89–98.

Update

Transportation Research Interdisciplinary Perspectives

Volume 9, Issue , March 2021, Page

DOI: <https://doi.org/10.1016/j.trip.2021.100308>



Erratum

Erratum regarding missing Declaration of Competing Interest statements in previously published articles



Declaration of Competing Interest statements were not included in the published version of the following articles that appeared in previous issues of "Transportation Research Interdisciplinary Perspectives".

The appropriate Declaration/Competing Interest statements, provided by the Authors, are included below.

1. "Unsafe motorization: A clog in the wheels of sustainable transportation" [Transportation Research Interdisciplinary Perspectives, 2020; Volume 6: 100153] <https://doi.org/10.1016/j.trip.2020.100153>

Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.

2. "An exploration of policy knowledge-seeking on high-volume, low-carbon transport: findings from expert interviews in selected African and South-Asian countries" [Transportation Research Interdisciplinary Perspectives, 2020; Volume 5: 100117] <https://doi.org/10.1016/j.trip.2020.100117>

Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.

3. "Agent-based vulnerability assessment at airport security checkpoints: A case study on security operator behavior" [Transportation Research Interdisciplinary Perspectives, 2020; Volume 5: 100139] <https://doi.org/10.1016/j.trip.2020.100139>

Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.

4. "The effect of COVID-19 and subsequent social distancing on travel behavior" [Transportation Research Interdisciplinary Perspectives, 2020; Volume 5: 100121] <https://doi.org/10.1016/j.trip.2020.100121>

Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.

5. "Air quality and fossil fuel driven transportation in the Metropolitan Area of São Paulo" [Transportation Research Interdisciplinary Perspectives, 2020; Volume 5: 100137] <https://doi.org/10.1016/j.trip.2020.100137>

Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.

6. "The ultimate smart mobility combination for sustainable transport? A case study on shared electric automated mobility initiatives in the Netherlands" [Transportation Research Interdisciplinary Perspectives, 2020; Volume 5: 100129] <https://doi.org/10.1016/j.trip.2020.100129>

Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.

7. "A framework for end-to-end deep learning-based anomaly detection in transportation networks" [Transportation Research Interdisciplinary Perspectives, 2020; Volume 5: 100112] <https://doi.org/10.1016/j.trip.2020.100112>

Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.

8. "Ethical decision making behind the wheel – A driving simulator study" [Transportation Research Interdisciplinary Perspectives, 2020; Volume 5: 100147] <https://doi.org/10.1016/j.trip.2020.100147>

Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.

9. "Population size and transport company efficiency – Evidence from Czech Republic" [Transportation Research Interdisciplinary Perspectives, 2020; Volume 6: 100145] <https://doi.org/10.1016/j.trip.2020.100145>

Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.

10. "Toward sustainable travel: An analysis of campus bikeshare use" [Transportation Research Interdisciplinary Perspectives, 2020; Volume 6: 100162] <https://doi.org/10.1016/j.trip.2020.100162>

Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.

11. "Responsible Transport: A post-COVID agenda for transport policy and practice" [Transportation Research Interdisciplinary Perspectives, 2020; Volume 6: 100151] <https://doi.org/10.1016/j.trip.2020.100151>

Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.

DOI of original article: <https://doi.org/10.1016/j.trip.2020.100147>. <https://doi.org/10.1016/j.trip.2020.100144>. <https://doi.org/10.1016/j.trip.2020.100153>. <https://doi.org/10.1016/j.trip.2020.100142>. <https://doi.org/10.1016/j.trip.2019.100060>. <https://doi.org/10.1016/j.trip.2020.100123>. <https://doi.org/10.1016/j.trip.2020.100139>. <https://doi.org/10.1016/j.trip.2020.100112>. <https://doi.org/10.1016/j.trip.2020.100154>. <https://doi.org/10.1016/j.trip.2020.100117>. <https://doi.org/10.1016/j.trip.2020.100121>. <https://doi.org/10.1016/j.trip.2020.100145>. <https://doi.org/10.1016/j.trip.2020.100162>. <https://doi.org/10.1016/j.trip.2020.100151>. <https://doi.org/10.1016/j.trip.2020.100118>. <https://doi.org/10.1016/j.trip.2020.100109>. <https://doi.org/10.1016/j.trip.2020.100159>. <https://doi.org/10.1016/j.trip.2020.100137>. <https://doi.org/10.1016/j.trip.2020.100129>. <https://doi.org/10.1016/j.trip.2020.100150>. <https://doi.org/10.1016/j.trip.2020.100110>. <https://doi.org/10.1016/j.trip.2020.100136>.

<https://doi.org/10.1016/j.trip.2021.100308>

Available online 15 February 2021

2590-1982/© 2021 Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

12. “Distracted by “distracted pedestrians”?” [Transportation Research Interdisciplinary Perspectives, 2020; Volume 5: 100118] <https://doi.org/10.1016/j.trip.2020.100118>
Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.
13. “The cruise industry and the COVID-19 outbreak” [Transportation Research Interdisciplinary Perspectives, 2020; Volume 5: 100136] <https://doi.org/10.1016/j.trip.2020.100136>
Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.
14. “Estimating bus passenger volume based on a Wi-Fi scanner survey” [Transportation Research Interdisciplinary Perspectives, 2020; Volume 6: 100142] <https://doi.org/10.1016/j.trip.2020.100142>
Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.
15. “Impact of COVID-19 on transportation in Lagos, Nigeria” [Transportation Research Interdisciplinary Perspectives, 2020; Volume 6: 100154] <https://doi.org/10.1016/j.trip.2020.100154>
Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.
16. “Rail factor and realism of the unconscious” [Transportation Research Interdisciplinary Perspectives, 2020; Volume 6: 100144] <https://doi.org/10.1016/j.trip.2020.100144>
Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.
17. “How COVID-19 and the Dutch ‘intelligent lockdown’ change activities, work and travel behaviour: Evidence from longitudinal data in the Netherlands” [Transportation Research Interdisciplinary Perspectives, 2020; Volume 6: 100150] <https://doi.org/10.1016/j.trip.2020.100150>
Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.
18. “Interprofessional collaboration to promote transportation equity for environmental justice populations: A mixed methods study of civil engineers, transportation planners, and social workers’ perspectives” [Transportation Research Interdisciplinary Perspectives, 2020; Volume 5: 100110] <https://doi.org/10.1016/j.trip.2020.100110>
Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.
19. “Framing systemic traffic violence: Media coverage of Dutch traffic crashes” [Transportation Research Interdisciplinary Perspectives, 2020; Volume 5: 100109] <https://doi.org/10.1016/j.trip.2020.100109>
Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.
20. “Transit in flex: Examining service fragmentation of app-based, on-demand transit services in Texas” [Transportation Research Interdisciplinary Perspectives, 2019; Volume 5: 100060] <https://doi.org/10.1016/j.trip.2019.100060>
Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.
21. “COVID-19 and airline employment: Insights from historical uncertainty shocks to the industry” [Transportation Research Interdisciplinary Perspectives, 2020; Volume 5: 100123] <https://doi.org/10.1016/j.trip.2020.100123>
Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.
22. “Inflection point: The future of subcontracting in the petroleum industry” [Transportation Research Interdisciplinary Perspectives, 2020; Volume 6: 100159] <https://doi.org/10.1016/j.trip.2020.100159>
Declaration of competing interest: The authors were contacted after publication to request a Declaration of Interest statement.