SCIENCE CHINA
Information Sciences

• LETTER •

March 2021, Vol. 64 139202:1–139202:3
https://doi.org/10.1007/s11432-018-9650-9

Deep amended COPERT model for regional vehicle emission prediction

Zhenyi XU[1], Yu KANG[1,2*], Yang CAO[1] & Zerui LI[1]

[1]*Department of Automation, University of Science and Technology of China, Hefei* 230026*, China;*
[2]*Key Laboratory of Technology in GeoSpatial Information Processing and Application System, Chinese Academy of Sciences, Beijing* 100192*, China*

Received 14 August 2018/Accepted 31 October 2018/Published online 13 April 2020

Citation Xu Z Y, Kang Y, Cao Y, et al. Deep amended COPERT model for regional vehicle emission prediction. Sci China Inf Sci, 2021, 64(3): 139202, https://doi.org/10.1007/s11432-018-9650-9

Dear editor,

The existing methods for regional vehicle emission prediction can be roughly categorized into the classes of classical dispersion models and satellite remote sensing models. Gaussian plume models, operational street canyon models and computational fluid dynamics are the classical dispersion models. The mobile source emission factor (MOBILE) and computer programme to calculate emissions from road transport (COPERT) models that have been developed in USA and Europe respectively are the most frequently used emission factor models [1]. These models [2, 3] are usually complex models that consider the meteorology, road network geometry, geographical locations, traffic volumes, and emission factors based on a several empirical assumptions and parameters that may not be applicable to all the regions in a city [4]. The aforementioned parameters are usually considered to be difficult and expensive to obtain, and the results that are generated using these parameters may be inaccurate. Further, satellite remote sensing of the surface air pollution has been extensively investigated during previous decades [5], and can be considered to be a top-down method. However, such approaches are extremely influenced by the presence of clouds and are considered to be sensitive to other environmental factors such as humidity, temperature, pressure and geographical location [6, 7].

In addition, forecasting the vehicle emission in each region of a city is very challenging because of the insufficient number of remote sensing vehicle measurement stations available in a city for the expensive cost of building and maintaining measurement equipments. And urban vehicle emission varies by locations non-linearly and depends on several complex external factors, including road networks, meteorology, traffic, green land ratio and the type of living function area.

To address the aforementioned issues, we predict the regional vehicle emission using a data-driven method in this study, by employing a variety of datasets, including the remote sensing records, meteorological data, traffic data, data from road networks and points of interest (POIs). Further, we propose an amended COPERT prediction model to address the data sparsity and spatial heterogeneity that includes complex external factors for generating spatiotemporal emission data to pre-train the deep spatiotemporal network (DeepST); additionally, we formulate the region emission prediction as a spatiotemporal sequence forecasting problem that can be solved by constructing a deep learning framework. For modeling the spatiotemporal relations, DeepST can be employed to collectively predict the vehicle emission in every region. The effectiveness of the proposed method can be demonstrated by comparing it with several baseline methods using a real-world dataset.

Problem Formulation. For road vehicle emission calculation, there are several emission models that employ different approaches for emission computation. The choice of a modeling approach is considered to be dependent on the purpose of computation, such as macro emission inventory and micro emission factor.

The amount of pollutants that are emitted by a single vehicle can be calculated as the function of the travel speed v (km/h). Further, the general function can be written as

$$\mathrm{EF} = \frac{a + cv + ev^2}{1 + bv + dv^2}. \tag{1}$$

The parameters that are required for calculating different kinds of emissions and the amount of gas consumption are provided in Appendix A. Further, the overall emission on a certain road r can be given as

$$E = \mathrm{EF} \times R_{N_a} \times R_n \times R_{\mathrm{len}}, \tag{2}$$

where R denotes a given road segment, R_{N_a} denotes the traffic volume, R_n denotes the number of road lines, and R_{len} denotes the length of the road segment. By introducing various external factors, such as POIs, meteorology, road networks, and traffic volume, into the COPERT model [8], a deep learning framework can be designed to amend the COPERT model, thereby obtaining historical emission observations. At this time, the historical emission observation

* Corresponding author (email: kangduyu@ustc.edu.cn)

sequence X_k $(k = 0, 1, \dots, t-1)$ and external factors that are required to predict X_t should be provided.

Methodology. We propose an amended COPERT prediction model based on the deep spatiotemporal network. First, we employ a two-layer autoencoder to extract the external features from multi-source heterogeneous meteorology data, road networks, traffic and POIs. We further obtain the amended COPERT model by introducing external features that can be used to generate spatiotemporal data for the DeepST model. By retraining the DeepST model using the available real station measurements, a deep amended COPERT model can be obtained to predict the emission for the region.

The amended COPERT model. Because the COPERT model does not consider the external factors, such as temperature, wind, and POIs, we introduce an error term $\triangle E$ to amend (2), so as to ensure usage of the available meteorological and POI data. The error term denotes the uncertainties that can be attributed to external factors and that will be fixed by our deep spatiotemporal learning approach in the following. Further, the amended COPERT model can be written as

$$E = \beta_{\text{poi}} \times \ln t \times e^{-w} \times \text{EF} \times R_{N_a} \times R_n \times R_{\text{len}} + \triangle E(t, w, \text{POIs}), \tag{3}$$

where

$$\beta_{\text{poi}} = \begin{cases} \dfrac{\text{Num}_{\text{ng}}}{\text{Num}_{\text{ng}} + \text{Num}_g}, & \text{Num}_{\text{ng}} \neq 0, \\ 0.5, & \text{otherwise}, \end{cases}$$

Num_g denotes the number of green area POIs, Num_{ng} denotes the number of non-green area POIs, t denotes the temperature, and w denotes the wind speed.

Deep spatiotemporal network. To pre-train the DeepST model, we initially construct the training examples based on the original historical sequence records. Further, the DeepST model is trained by back-propagation and Adam algorithm [9]. The training procedure algorithm is given in Algorithm 1. The details of DeepST model is clarified in Appendix E.

Algorithm 1

• Prepare the required input, including the historical observations $\{X_0, X_1, \dots, X_{n-1}\}$, external features $\{E_0, E_1, \dots, E_{n-1}\}$, lengths of closeness sequences l_c, lengths of period sequences l_p, lengths of trend sequences l_s, period interval p, and trend interval s.
• Create training instances $D \leftarrow \emptyset$.
• **For** all available time interval t $(1 \leqslant t \leqslant n-1)$:

$$H_c = [X_{t-l_c}, X_{t-(l_c-1)}, \dots, X_{t-1}],$$

$$H_p = [X_{t-l_p \cdot p}, X_{t-(l_p-1)\cdot p}, \dots, X_{t-p}],$$

$$H_s = [X_{t-l_s \cdot s}, X_{t-(l_s-1)\cdot s}, \dots, X_{t-s}],$$

and put a training instance $(\{H_c, H_p, H_s, E_t\}, X_t)$ into D.
• Initialize the optimal parameter θ in the DeepST model.
• Randomly select a batch of instances D_b from D.
• Find θ by minimizing the equation $L(\theta) = \|\hat{X}_t - X_t\|_2^2$ using D_b.
• Output the learned DeepST model M.

After the pre-training is completed, we obtain the pre-trained DeepST model M; further, we can use the available station records to retrain the model for determining the error term $\triangle E$ to amend the COPRET. Finally, an amended DeepST model is obtained for performing regional emission prediction. The process of prediction using the learned model M is presented in Algorithm 2.

Algorithm 2

• Prepare the required input, including the trained DeepST model M, number of look-ahead prediction steps k, historical sequence observations $\{X_0, X_1, \dots, X_{n-1}\}$, external features $\{E_0, E_1, \dots, E_{n-1}\}$, lengths of closeness sequences l_c, lengths of period sequences l_p, lengths of trend sequences l_s, period interval p, and trend interval s.
• $\mathbb{X} \leftarrow \{X_0, X_1, \dots, X_{n-1}\}$.
• **For** $t = n$ **to** $n + k - 1$:

$$H_c = [X_{t-l_c}, X_{t-(l_c-1)}, \dots, X_{t-1}],$$

$$H_p = [X_{t-l_p \cdot p}, X_{t-(l_p-1)\cdot p}, \dots, X_{t-p}],$$

$$H_s = [X_{t-l_s \cdot s}, X_{t-(l_s-1)\cdot s}, \dots, X_{t-s}],$$

$$\hat{X}_t \leftarrow M(H_c, H_p, H_s, E_t),$$

and insert the prediction instance $\hat{X}_t$ into $\mathbb{X}$.
• Output $\{\hat{X}_n, \hat{X}_{n+1}, \dots, \hat{X}_{n+k-1}\}$.

Experiments. The proposed method is implemented on a personal computer, with the specifications being provided in Appendix B. A detailed description of the datasets has been provided in the supplementary material. In this experiment, we compare CO and NO_x with two other models based on the records of Hefei city, as depicted in Appendix C. To the best of our knowledge, the regional emission prediction problem exhibits inherent spatiotemporal dependencies, because our model simultaneously considers temporal and spatial dependencies which is superior to the artificial neural network (ANN) and recurrent neural network (RNN) models. We provide the variants of DeepST using different external factors, presented in Appendix D, and observe that the DeepST model when combined with the traffic, wind and temperature factors are better than the other baselines models. This significantly improves the accuracy and denotes that our model can learn the map function between the complex external factors and the emission if we can obtain the relevant external factors .

Conclusion. We proposed a deep amended COPERT prediction model to address data sparsity and spatial heterogeneity; further, we formulated the regional emission prediction as a spatiotemporal sequence forecasting problem that can be solved by constructing a deep learning framework. The effectiveness of the proposed method was demonstrated through comparison with several baseline methods using real-world data.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant Nos. 61725304, 61673361).

Supporting information Appendixes A–E. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Smit R, Ntziachristos L, Boulter P. Validation of road vehicle and traffic emission models – a review and meta-analysis. Atmos Environ, 2010, 44: 2943–2953

2 Chang X, Chen B Y, Li Q, et al. Estimating real-time traffic carbon dioxide emissions based on intelligent transportation system technologies. IEEE Trans Intell Transp Syst, 2013, 14: 469–479

3 Zito P, Chen H B, Bell M C. Predicting real-time roadside CO and NO_2 concentrations using neural networks. IEEE Trans Intell Transp Syst, 2008, 9: 514–522

4 Vardoulakis S, Fisher B E A, Pericleous K, et al. Modelling air quality in street canyons: a review. Atmos Environ, 2003, 37: 155–182

5 Martin R V. Satellite remote sensing of surface air quality. Atmos Environ, 2008, 42: 7823–7843

6 van Donkelaar A, Martin R V, Park R J. Estimating ground-level $PM_{2.5}$ using aerosol optical depth determined from satellite remote sensing. J Geophys Res, 2006, 111: 21201

7 Xie Z W, Zeng Z, Zhou G Y, et al. Topic enhanced deep structured semantic models for knowledge base question answering. Sci China Inf Sci, 2017, 60: 110103

8 Ntziachristos L, Samaras Z. COPERT II: Computer Programme to Calculate Emissions from Road Transport: User Manual. 1997

9 Kingma D P, Ba J. Adam: a method for stochastic optimization. 2014. ArXiv:1412.6980