Transportation Research Part D 63 (2018) 236–243

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Use of Bayesian inference method to model vehicular air pollution in local urban areas

A. Orun*, D. Elizondo, E. Goodyer, D. Paluszczyszyn

School of Computer Science and Informatics, De Montfort University, Leicester LE1 9BH, United Kingdom

ARTICLE INFO

Keywords:
Vehicular air pollution
Artificial intelligence
Predictive modelling
GIS

ABSTRACT

Traffic Related Air Pollution (*TRAP*) studies are usually investigated using different categories such as air pollution exposure for health impacts, urban transportation network design to mitigate pollution, environmental impacts of pollution, etc. All of these subfields often rely on a robust air pollution model, which also necessitates an accurate prediction of future pollutants. As is widely accepted by the heath authorities, *TRAP* is considered to be the major health issue in urban areas, and it is difficult to keep pollution at harmless levels if the time sequenced dynamic pollution and traffic parameters are not identified and modelled efficiently. In our work here, artificial intelligence techniques, such as Bayesian Networks with an optimized configuration, are used to deliver a probabilistic traffic data analysis and predictive modelling for air pollution (SO_2, NO_2 and CO) at very local scale of an urban region with up to 85% accuracy. The main challenge for traditional data analysis is a lack of capability to reveal the hidden links between distant data attributes (e.g. pollution sources, dynamic traffic parameters, etc.), whereas some subtle effects of these parameters or events may play an important role in pollution on a long-term basis. This study focuses on the optimisation of Bayesian Networks to unveil hidden links and to increase the prediction accuracy of *TRAP* considering its further association with a predictive GIS system.

1. Introduction

The term air pollution is linked to harmful substances in the air such as gases, solid particles or liquid droplets. The more specifically, traffic related air pollutants may be categorised as primary and secondary pollutants where the first category incudes carbon monoxide, sulphur dioxide, nitrogen oxides (Alvarez-Vazquez et al., 2017) and also Hydrocarbons, particulate matter, mobile-source air toxics such as lead, benzene, and formaldehyde (Health Effects Institute, 2010) whereas the second category covers ozone and the other minor pollutants. The pollutant emission measures are usually described as the weight of pollutant divided by a unit weight, volume, distance or duration of the pollution activity (US Environmental Pollution Agency, 1995). The air pollution is automatically measured by Air quality stations located at the critical points of urban areas where population or traffic flow is at certain density.

Nowadays with the increasing population of different vehicle types and by inadequate traditional transport or traffic system designs, air pollution has become one of the key issues to be solved urgently for urban areas. Traffic related air pollution (*TRAP*), which contains harmful chemicals, is a major threat for cities. In dense urban areas, vehicle emissions may be responsible for 90–95% of carbon monoxide and 60–70% of nitrogen oxides within the atmosphere (Schwela, 2000). Recent epidemiological studies suggest

* Corresponding author.
E-mail address: aorun@dmu.ac.uk (A. Orun).

https://doi.org/10.1016/j.trd.2018.05.009

1361-9209/ © 2018 Elsevier Ltd. All rights reserved.

that at present air pollution results in an average 7-month reduction in life expectancy and costs UK society up to £20 billion per year (HCEA Committee, 2011).

The immediate solution is not easy if the time sequenced dynamic pollution and traffic system parameters are not properly identified and modelled. Particularly multidisciplinary areas such as artificial intelligence (AI) methods (e.g. data mining, inference methods, etc.), state-of-art instrumentation, supercomputing facilities, distributed sensors, etc. would be expected to bring the most promising solutions to the problem. In our work here an artificial intelligence technique, Bayesian Networks, are used for a probabilistic data analysis whose performance has already been proven by our previous works (Orun and Aydin, 2010; Orun, 2004). One of the traditional common issues of manual data analysis is the lack of visibility of the hidden links between distant and the least correlated data attributes (e.g. pollution sources, dynamic traffic parameters, geographic location characteristics, etc.); whereas some subtle effects of these parameters or specific combinations of them may play an important role in traffic related air pollution on a long-term basis. The proposed work here targets an optimised Bayesian model to unveil such hidden links between the *TRAP* factors and parameters.

Many works have previously investigated air pollution. The most similar methodology was used with Bayesian classifier by Corani and Scanagatta (2016), but only for the prediction of $PM_{2.5}$ (Atmospheric aerosol particles) pollutant and O_3 (Ozone gas). In another work Karatzas and Kaltsatos introduce a computational intelligence method (Karatzas and Kaltsatos, 2007) for air pollution modelling by which the environmental system is simulated. Their work was also at a large geographic scale for a city area. Zhu et al. (2015) investigate traffic-related air pollution in a street canyon by utilising a genetic algorithm-back propagation artificial neural network but even though it introduces an AI-based approach, it is only based on a single pollutant parameter, nitrogen dioxide (NO_2), rather than focusing on a multi parameter solution. In (Passow et al. 2012) authors also used NO_2 used as a tracer for pollution but included additional factors derived from meteorological data, traffic data and earth observation data to create framework for near-real time traffic management and air quality control. In contrast, the method introduced in this paper is dedicated to long-term forecast to support design and re-assessment of air quality models in local urban areas. The method proposed here will also include several diverse types of parameters (pollutant, environmental, etc.), processed in an interactive form, to identify subtle connection between them.

On the other hand, geographic information systems (GIS) (Orun, 1993) are one of the major tools that could also potentially be used for the transport planning (Niemeier and Beard, 1993). One of the earlier works (Thonga and Wongb, 1997) introduced a specific database design for use by GIS for urban transport planning. In the work, even though some “predictive” techniques (called what-if) were suggested, their implementation was not considered within the proposed system. Another statistical predictive model based on a “distance decay regression” approach was introduced by Jason et al. (2009). Even though the method effectively used statistical algorithms with high prediction accuracy, it was limited with NO_x gases and no GIS utility was used or evaluated for a pollution map formation. The majority of past studies Jason et al. (2009), Matthias et al. (2006), Zhu et al. (2015), and Shekarrizfard et al. (2017) were limited to NO_x pollutants, whereas in our work the additional SO_2 and CO based pollutions were also studied.

Only a few earlier works combined the prediction methods with GIS for a pollution analysis. A direct estimation of traffic related air pollution was made by a GIS based component-oriented integrated system which was developed by Rebolj and Sturm (1999). In the paper, the generic outline of proposed system was introduced. However, the presentation did not exhibit the numerical test results in details other than a single output. The current statistical traffic-related air pollution models (Gulliver and Briggs, 2005) used for traffic design may cause several problems such as, Available data may not provide accurate pollution information for the future if processed by inappropriate linear predictive models. Such inappropriate modelling may later cause bulky re-construction and development for road network amendments.

Within this work we overcome the above issues by deploying a predictive modelling based GIS system in which AI technique, Learning Bayesian Networks, is used for a new target layer generation.

This paper is organised as follow; The method and materials used here are introduced in Section 2, with a presentation of a limited data sample set. In the subsections Bayesian Networks and predictive Bayesian model are also introduced by the presentation of a flow chart and pseudo code of the proposed method. The results of model accuracy are finally represented in Section 3 as results and discussion followed by the conclusion section.

2. Methods and materials

2.1. Data set specifications

The restricted data set consists of weekly recordings of traffic flows (at local data collection stations), air pollution values (e.g. SO_2, NO_2, CO), local temperature readings, wind records, air pressure, rainfall and global radiation values within Leicester City local urban areas for the year 2012. The local regions selected in this work are Newark and Aunsite. The whole data set was utilised for Bayesian Network (BN) construction as seen in Fig. 3, where the abbreviations “st” refers to traffic data collection stations in the Leicester city area. The traffic flow data were collected over the 56 station locations with additional 9 parameters including pollution types, temperature, global radiation, air pressure, rainfall, wind speed and wind direction.

Parameter units in Table 1: (NO_2, SO_2) = parts per billion, CO = parts per million, Temperature = °C, Global Radiation = W/m^2, Air pressure = mbar, Wind direction = Degree.M, Wind speed = m/s, Rainfall = mm/h, traffic flow (st_i) = number of vehicles/h.

The collected raw data need additional processing in which different fractions of data segments with different data types were re-arranged so they can be integrated within this work. This inevitably caused a limitation of data volume and relatively limited training of the BN system at low efficiency.

Download English Version:

https://daneshyari.com/en/article/7498370

Download Persian Version:

https://daneshyari.com/article/7498370

Daneshyari.com