TPM- Transport Practitioner's Meeting Conference, 28-29 June 2017, Nottingham

An Intelligent traffic network optimisation by use of Bayesian inference methods to combat air pollution in local urban areas

D. Elizondo and A.B Orun
School of Computer Science and Informatics
De Montfort University, Leicester LE1 9BH
E-mail: elizondo@dmu.ac.uk, Phone: 0116-2078471

Keywords *: traffic network design, optimisation, air pollution, Bayesian networks.*

Abstract

Traffic flow related air pollution is one of the major problems in urban areas, and is often difficult to avoid it if the time sequenced dynamic pollution and traffic parameters are not identified and modelled efficiently. In our introduced work here, an artificial intelligence technique such as Bayesian networks are used for a robust traffic data analysis and modelling. The most common challenge in traditional data analysis is a lack of capability of unveiling the hidden links between the distant data attributes (e.g. pollution sources, dynamic traffic parameters, geographic location characteristics, etc.), whereas some subtle effects of these parameters or events may play an important role in pollution on a long term basis.

1 Introduction

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 major issues to be solved urgently for the urban areas. Traffic related air pollution is a major thread for cities that contains harmful chemicals. The straightaway solutions may seem to be not easy tasks if the time sequenced dynamic pollution and traffic system parameters are not properly identified and modelled by a novel approach. Particularly multidisciplinary areas such as artificial intelligence methods (e.g. data mining, inference methods, etc.), state-of-art instrumentation, super computing, distributed sensors, etc. would be expected to bring most promising solutions to the problem. In our introduced work here an artificial intelligence technique such as Bayesian networks are used for a robust data analysis whose performance was already proven by our previous works [1][2] . One of the traditional common issues of a manual data analysis is the lack of unveiling the hidden links between the distant and least correlated data attributes (e.g. pollution sources, dynamic traffic parameters, geographic location characteristics, etc.) whereas some subtle effects of these parameters or events may play an important role in traffic related air pollution on a long term basis. Several works have been done previously to investigate air pollution. Olvera-Garcia uses Fuzzy inference system (Olvera-Garcia, et al. 2016) for air quality assessment by generating an air quality index. But their generic "non-traffic" based study covers very large area (Mexico city) of air pollution rather than a local urban regions which does not bring region-specific solutions. In the other work Karatzas and Kaltsatos introduce a computational intelligence method (Karatzas and Kaltsatos, 2007) for an air pollution modelling by which the environmental system is simulated. Their work was also at larger geographic scale for a city area. Zhu et al. (2015) investigate a traffic-related air pollution in street canyon by utilizing genetic algorithm-back propagation artificial neural network but based only on a single pollutant parameter (NO_2) rather than multi parameter. whereas in our work several parameters (pollutant, environmental, etc.) are processed in an interactive form.

2 Methods and materials

2.1 Data set specifications

The restricted data set consists of a weekly recordings of traffic flow, 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 whole data set was utilized for Bayesian Network construction as seen in Figure 1, 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 points and additional 9 parameters including pollution types, temperature, wind speed and direction, etc.

hours	date	NO2_aunsite	CO_Newark	SO2_Newark	Temp	global_radi	Air_pressure	Wind_directic	Wind_speed	Reinfall	st1
8	120102	7.06503	-0.183	0.574682	2.235	0	994	199.4	5.09499	0	60
9	120102	9.58148	-0.0978333	0.395104	2.455	0	995	182.9	3.945	0	107
10	120102	10.5792	-0.164833	0.0854242	3.556	0.044	996	197.2	5.30299	0	184
11	120102	8.8035	0.0105	0.137453	4.22	0	996	197.7	4.48299	0	255
12	120102	8.74866	0.0353333	1.1158	5.261	0.224	996	208.7	5.689	0	298
13	120102	8.41095	0.0838332	4.62417	5.713	0.216	997	208.6	5.42799	0	323

Table 1. Partial display of the traffic data set. The whole set contains total number of 60 stations (only one station “st1” is shown). Station values correspond to traffic flow at the specific city locations.

2.2 . Data analysis by Bayesian networks

In general terms, Bayesian networks are called Casual Probabilistic Networks and very useful instrument which achieves an efficient knowledge representation and reasoning. They are also capable of generating very accurate classification results under uncertainty where the data set may include many uncertain conditions (Koski & Noble, 2009). The Bayesian networks graphically encode and represent the conditional independence (CI) relationships among a set of data (Orun, 2004). In this work, a learning Bayesian network software tool (PowerConstructor™) is used for the analysis of air pollution, traffic and environmental data and the Bayesian inference to construct the network (Figure 1). The algorithm examines information flow between two highly related variables (attributes) from a data set and decides if these variables (e.g., traffic parameters, etc.) are independent or linked and it also investigates into how close the relationship between those variables is.

One of earlier examples in which a Bayesian approach for an analysis of air pollution data was introduced by Suggs and Curran (1983). In their work air pollution data and air quality standards were compared and a combination of pollution history with instrumental precision in a Bayesian probabilistic model was comprehensively discussed. Some of our previous works also focus on the different application fields of Bayesian inference method and classification process separately, which provide a useful guidance for this work (Orun, 2004; Orun & Aydin, 2010). In those works two different experiments were done by use of Bayesian network tool (PowerPredictor™) for the analysis of data produced by the real-time lab experiments.

PowerConstructor is a different tool than Bayesian classifier (Cheng et al., 2002). But both utilities use the Markov condition to obtain a collection of conditional independence statements from the networks (Pearl, 1988). One of the advantages of Bayesian networks over the other AI systems (e.g. Neural networks or Fuzzy Logic) is that, it exhibits direct and indirect links between the attributes which can be easily interpreted.

3 **Results and discussion**

As is seen in the Bayesian networks (Figure 1) which was built after an inference based data processing, the following conclusions would be drawn to interpret the links between the attributes (e.g. system parameters, data collection stations, etc.) in the network. Some of the conclusion made are as follows :

- Traffic data collection Station56 (shown as St56) has a central role as it has 7 connections with other stations (for vehicle flow data). This means that any structural change on Station 56 would have substantial effect on the other stations.
- Air pollution attributes have natural links with some parameters like: temperature, rainfall, CO_Newarke, air pressure, wind speed, NO_2. This will lead to a natural modelling to be associated with the main traffic air pollution model.
- CO in Newarke area has a link with traffic station 33., NO_2 Aunsite area has link with station 10, SO_2 Newarke area has link with station 59., it has also link with wind direction whose cause-effect relationships would need an efficient interpretation before a modelling.
- The stations (sti) for vehicle flow data collection have links with each others, which may give an idea about the interactions between the high density vehicle flow regions in the city. (particularly exhibited as hidden links)
- One of examples for a hidden interaction would be between temperature and CO pollution. Which may help in traffic planning to bring some restriction on CO emission (e.g. by speed reduction, etc.) in warm days of the year.

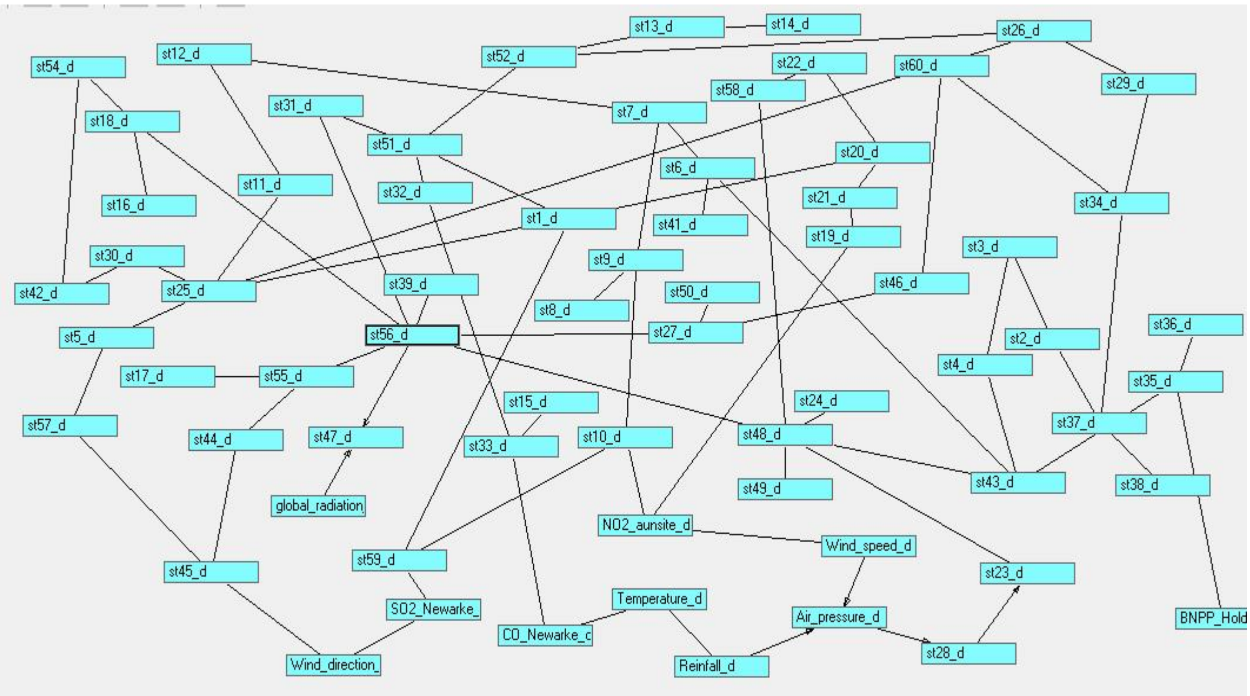

Figure 1. Established Bayesian network (BN) configuration after the use of inference tool called PowerConstructor™ to connect links between the data attributes.

3.1 Cause-effect connections

The data interpretation is the most challenging step among the data analysis which would otherwise not be possible to achieve manually by eye. In our work, local link analyses have been made as shown in Figures 2 – 6 where each was derived from the global Bayesian network configuration (grey boxes indicate the traffic data collection stations). In the BN utility attribute connection threshold "t" is set to 0.1 for maximum number of links. The dicretization method for data set values was selected as equal frequency.

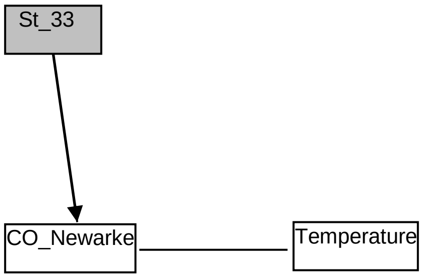

Figure 2. Traffic flow effect of Station33 on CO pollution in Newark area with an influence of temperature variations.

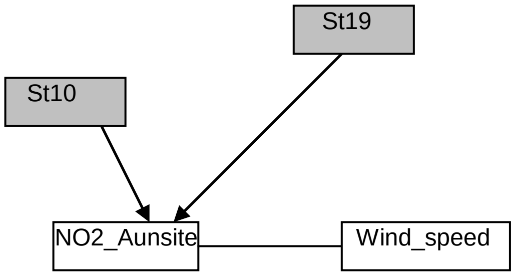

Figure 3. Traffic flow effects of stations St19 and St10 on NO_2 pollution in Aunsite area with the influence of wind speed variations

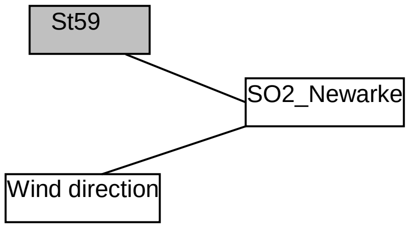

Figure 4. Traffic flow effects of stations St59 on SO_2 pollution in Newarke area with the influence of wind direction parameters

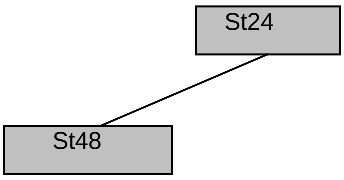

Figure 5. Traffic flow interaction between the stations St24 and St48 which may be taken into account for easing the traffic load of each one at traffic network design stage.

The examples of local cause-effect connections may provide useful tools for traffic design strategy to minimize the air pollution at the design stage. In example interpretation as shown in Figure 2, the CO pollution in Newarke area is caused by Station 33 as the temperature variation also seems to be an impact factor. If there was a direct link between the temperature and Station33, then the possible conclusion would be made that the temperature variation might be caused by high density of traffic congestions (But in this circumstance temperature is possibly the function of gas emission). In Figure 3, Traffic flow effects of stations St19 and St10 on NO_2 pollution in Aunsite area with the influence of wind speed variations, which concerns topographic characteristics of the location. Similarly in Figure 4, the pollution caused by Station 59 is under the influence of wind direction where it has to be taken into account during the traffic network design phase in regards to geographic location. In Figure 5 an interaction between the Stations provide a beneficial information for easing the traffic load on any of those station junctions by transferring its traffic flow to the other.

4 Conclusion

Within this work too many issues have been faced, particularly the data set construction was one of the major problem due to format incompatibility of different data collection sources (e.g. sensors, environmental data, traffic parameters, etc.). This would be resolved by automated format conversion algorithms which would otherwise be impossible to rearranging huge amount of data manually.

As far as other fundamental issue is concerned, an accurate and reliable modelling is always a big challenge for high parameter-interactive time-sequenced domains, like air pollution measures of a traffic area. Such a modelling issue would only be solved by state-of-art techniques such as Artificial Intelligence (data mining) in association with efficiently distributed low-cost sensor networks, etc. Our ultimate target within this work was an optimized method for a traffic air pollution modelling by which maximum desired reliability and accuracy would be obtained by use of feasible instrumentation and labour work at moderate cost (that is affordable by local governments). The method would be particularly useful at traffic network design stages where subtle parametric impacts would be more effective than expected on the environment and economy on the long term basis. The proposed model could be further extended beyond the current data restriction by additional parameters such as GIS (Geographic Information System) related factors including terrain topography information, natural environmental effects, etc. with an adequate system integration for further optimisation.

References

[1] Orun, A.B and Aydin, N. "Variable optimisation of medical image data by the learning Bayesian Network reasoning", 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'10), Buenos Aires, Argentina, 1st - 4th September, 2010.

[2] Orun, A.B. "Automated identification of man-made textural features on satellite imagery by Bayesian networks", Photogrammetric Engineering and Remote Sensing, Vol. 70, No. 2, February 2004

[3] Suggs, J.C. and T.C. Curran, An empirical Bayes method for comparing air pollution data to air quality standards. Atmospheric Environment (1967), Volume 17, Issue 4, 1983, Pages 837-841.

[4] Koski, T., & Noble, J. (2009). Bayesian networks: An introduction. UK: John Wiley and Sons.

[5] Cheng, J., D. Bell and W. Liu. (2002). "Learning Bayesian Networks from data : An efficient approach based on information theory", http://www.cs.ualberta.ca/~jcheng/lab.htm

[6] Pearl, J. (1988). Probabilistic reasoning in intelligence systems: Network of plausible inference. San Francisco, California: Morgan Kaufmann

[7] Olvera-Garcia, M.A., Carbajal-Hernandez, J.J., Sanches-Fernandez, L.P. and Hernandez-Bautista, I, Air Quality assessment using a weighted fuzzy inference system. Ecological Informatics, Vol.33, May 2016. pp. 57-74.

[8] Karatzas, K.D. and S. Kaltsatos, Air pollution modelling with the aid of computational intelligence methods in Thessaloniki, Greece. Simulation Modelling Practice and Theory, Vol.15, 10, November 2007, pp.1310-1319.

[9] Zhu, G., Zhang, P. Tshukudu, T., Yin, J., Fan, G. and Zheng, X. Forecasting traffic-related nitrogen oxides within a street canyon by combining a genetic algorithm-back propagation artificial neural network and parametric models. Atmospheric Pollution Research, Vol.6, 6, November 2015. pp.1087-1097.