Missing Piece Prevents Seeing the Big Picture

Dealing with missing values in air pollution and weather to identify environmental risk factors associated with diabetes

Mrs. Lakmini Wijesekara
School of Computer, Data and Mathematical Sciences,
Western Sydney University.
18570263@student.westernsydney.edu.au

Dr. Liwan Liyanage
School of Computer, Data and Mathematical Sciences,
Western Sydney University.
l.liyanage@westernsydney.edu.au

A/Prof. Michael O'Connor
School of Medicine,
Western Sydney University.
m.oconnor@westernsydney.edu.au

1. WHY?

Do you know that Diabetes is linked with air pollution and weather?

- There is a direct link between the global diabetes epidemic and climate change[1, 2, 3]
- Diabetes attributable to PM2·5 air pollution is significant [4, 5, 6]
- Diabetes incidence rate in the USA and prevalence of glucose intolerance worldwide increase with higher outdoor temperature [7]

Missing value problem of sensor data!

- Air pollution and weather data are collected through sensors
- It is very common to have missing values due to failures of sensors
- Large gaps with missing data is a problem where research is rarely carried out to deal with

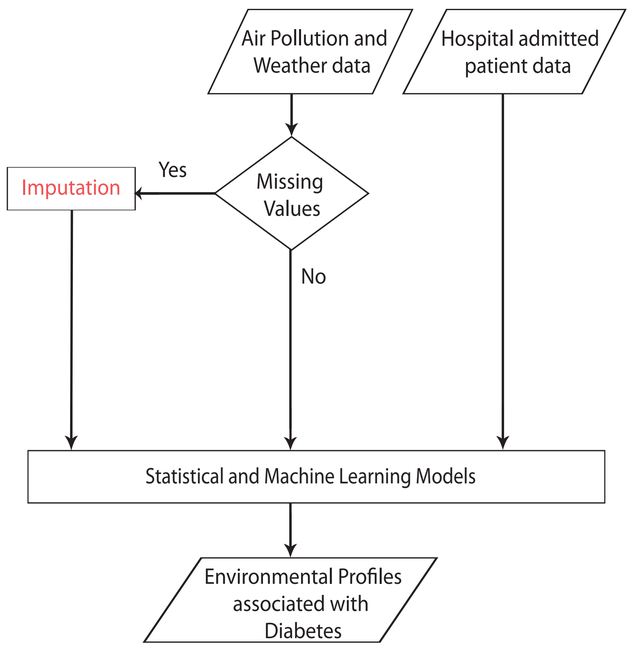
2. WHAT?

Two-fold objectives of our research!

- To develop a framework to deal with missing values in weather and pollution data, particularly for large gaps (Data Science Discipline)

- To identify vulnerability profiles of environmental conditions on diabetes (Application in Population Health)

3. HOW?



- Replacing a missing value by a plausible value is known as imputation

4. WHAT WE DID?

- We proposed an algorithm to impute large gaps using a bi-directional method based on regularized regression models [8,9]
- This method outperforms a set of baseline and well-established imputation methods (Figure 1)

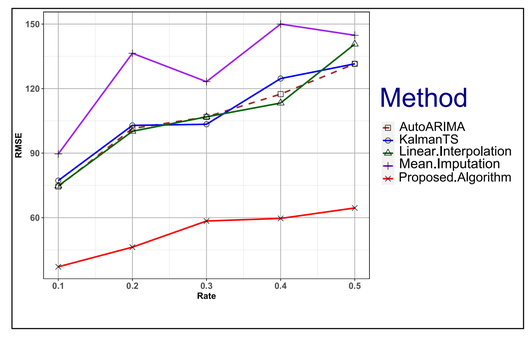


Figure 1: Root Mean Squared Error of the proposed method in comparison with a set of other methods for different propotions of missing values

- Developed a dashboard to visualize the patterns of missing values in NSW pollution and weather data (Figure 2)

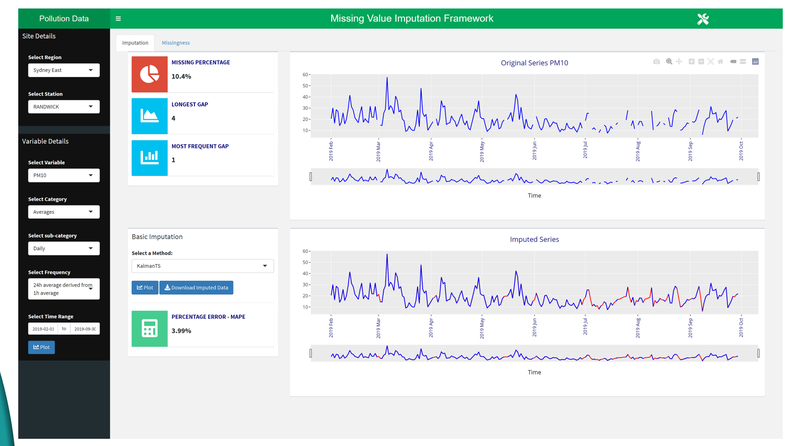


Figure 2: Dashboard to analyse missing values in NSW pollution data

- This provides facility to impute missing values using several methods and to download imputed data

5. SO WHAT?

- When using Environmental data captured by sensors having large gaps, this procedure will increase the accuracy of the results obtained through downstream analyses

- Such methodologies developed to deal with missing values could be extended to a wide range of applications influenced by environmental conditions

6. WHAT'S NEXT?

- Cleaned data will be linked with NSW Admitted Patient Data Collection from 2015 to 2020 and apply spatiotemporal modelling techniques to **identify vulnerability environmental profiles for diabetes** in NSW

- The findings of this research would be of great importance to policymakers to manage and mitigate the burden of diabetes

References

1. International Diabetes Federation, Diabetes and Climate Change Policy Report. 2012, International Diabetes Federation.
2. Colagiuri, R., Diabetes and climate change: different drums--same orchestra. Journal of Public Health Policy, 2013. 34(1): p. 165-9.
3. Cuschieri, S. and J. Calleja Agius, The interaction between diabetes and climate change – A review on the dual global phenomena. Early Human Development, 2021. 155: p. 105220.
4. Bowe, B., et al., The 2016 global and national burden of diabetes mellitus attributable to PM 2·5 air pollution. The Lancet Planetary Health, 2018. 2(7): p. e301-e312.
5. Pearson, J.F., et al., Association between Fine Particulate Matter and Diabetes Prevalence in the U.S. Diabetes Care, 2010. 33.
6. Dendup, T., et al., Environmental Risk Factors for Developing Type 2 Diabetes Mellitus: A Systematic Review. International Journal of Environmental Research and Public Health, 2018. 15(1): p. 78.
7. Blauw, L.L., et al., Diabetes incidence and glucose intolerance prevalence increase with higher outdoor temperature. BMJ Open Diabetes Res Care, 2017. 5(1): p. e000317.
8. Wijesekara, L. and L. Liyanage. Air quality data pre-processing: A novel algorithm to impute missing values in univariate time series. in 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI). 2021.
9. Wijesekara, L. and L. Liyanage. Imputing Large Gaps of High-resolution Environment Temperature. in 2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS). 2021.

Acknowledgment

We thank Accelerating Higher Education Expansion and Development (AHEAD) programme of Sri Lanka funded by World Bank for providing initial support and Western Sydney University Postgraduate Research Scholarship for providing continues support for this study. Also we thank Department of Planning and Environment, NSW Government for air pollution and weather data.

Contact us

Read more

WESTERN SYDNEY UNIVERSITY