Trabalho apresentado no XXXVIII CNMAC, Campinas - SP, 2018.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Categorical PCA and Multiple Correlation in the Study of the Incidence of Dengue Fever in Communities of Paraguay

Emilio Gerardo Sotto Riveros[1]
Polytechnic School, UNA, San Lorenzo, Paraguay
Santiago Gómez-Guerrero[2]
Polytechnic School, UNA, San Lorenzo, Paraguay
Christian Schaerer[3]
Center for Research in Mathematics, San Lorenzo, Paraguay

1 Introduction

The Principal Component Analysis (PCA) belongs to a group of multivariate statistical methods, and it is widely used as a descriptive technique in several fields to reduce the inherent complexity of having multiple variables. In public health observational studies, data of a mixed nature (discrete and continuous) are often found, as in [1]. Since the factors responsible for the epidemic and hemorrhagic varieties of dengue are complex and not yet fully understood, the Categorical Principal Components Analysis (CatPCA) emerges as an optimal tool to analyze the data collected. MSU, a novel measure of multiple correlation between variables, is also computed on the resulting PCA components to obtain a greater insight regarding the relevance of each variable.

2 Data and Methodology

In this work, data referring to suspected and confirmed cases of dengue were obtained from the records of the Ministry of Public Health of Paraguay [4], from different regions of the country and for a period of 5 years. These data were cleansed, standardized and grouped as part of the work in the COMIDENCO project [3]. On the other hand, climatological records were downloaded from Weather Underground [5], a meteorological service that provides information in real time through the Internet.

For the creation of matrices containing epidemiological and climatological variables with possible relationships over time, auxiliary files were generated with SQL queries, that allowed to organize the registers in a way to explore and compare different date

[1]emiliosotto@gmail.com
[2]sgomezpy@gmail.com
[3]chris.schaerer@cima.org.py

ranges of climatological events (rainfall, wind, temperature, humidity, pressure) with the appearance of symptoms of the disease (fever, headache, muscle aches) in order to identify which differences of days determine greater effects.

The CatPCA allows to scale the variables at different levels, and the categorical variables are optimally quantified in the specified dimensionality. As a result, non-linear relationships between the variables can be modeled.

Finally, the *Multivariate Symmetrical Uncertainty* (MSU) entropy-based measure, given by the total correlation between n categorical variables normalized by the sum of their individual entropies, is employed to assess the relative sizes of the variances explained by each group of variables obtained in the largest principal components.

3 Conclusions

Iteratively combining PCA results with MSU measurements is effective in supporting efforts to reach a deeper comprehension of factors that have influence over the behavior of cyclic epidemics such as dengue fever. In preliminary results, climatic variables as well as the blood NS1(AG) protein and the presence of IgG and IgM antibodies display possibly substantial association with the incidence of dengue fever in infected survival and death cases. Hence, employing two or more mathematical techniques together can be proven as valuable and cost-effective in the development of better prevention and control programs.

Acknowledgements. The authors thank PRONII-CONACyT-Paraguay. This work is supported by CIMA through research grant PINV15-706 from CONACyT.

References

[1] S. A. Ahmed, J. S. Siddiqi, S. Quaiser. *Principal Component Analysis to Explore Climatic Variability that Facilitates the Emergence of Dengue Outbreak in Karachi.* Pakistan Journal of Meteorology Vol. 11, Issue 21, 2014.

[2] G. Sosa-Cabrera, M. García-Torres, S. Gómez, C. Schaerer and F. Divina. *Understanding a Version of MSU to assist in Feature Selection.* Proceedings of the 4th Conference of Computational Interdisciplinary Science CCIS, Sao José dos Campos, 2016.

[3] S. Gómez, C. Schaerer, A. Rojas de Arias, H. Estigarribia, J. Mello. *Construcción de un modelo de incidencia de dengue aplicado a comunidades de Paraguay.* Proceedings of Segundo Encuentro de Investigadores de la Sociedad Científica del Paraguay, Asunción, 2017.

[4] Dirección General de Vigilancia Sanitaria. Boletín Epidemiológico Semanal. Available in http://www.vigisalud.gov.py/boletin_epidemiologico. Paraguay, 2009-2014.

[5] Weather Underground, a commercial weather service providing real-time weather information via the Internet. Available in https://espanol.wunderground.com/