Application of SOM based clustering on aggregated epidemiological data based on Nomenclature of Units for Territorial Statistics (NUTS)

Mehrdad A. Mizani [a], M. Oya Çınar [b], S. Can Yücebaş [a]

[a] *Informatics Institute, Medical Informatics Dpt, METU, Ankara, Turkey*

[b] *Biomedical Engineering Dpt, Baskent University, Ankara, Turkey*

Abstract: There is a huge amount of health data in Turkey that can be used for epidemiology, public health, and other non-clinical purposes. This data is of tremendous value when analyzed using data mining techniques revealing patterns that are not evident in conventional data analysis techniques. These hidden patterns are of value for epidemiological studies by revealing distributions in huge sets of data. Data mining techniques are most suitable for datasets with high levels of quantity and details. However accessing the raw, detailed, or individual based data is difficult, if not impossible**,** in Turkey due to privacy concerns or the lack of detailed nationwide regulations and procedures for data disclosure. We applied the SOM based clustering technique to the aggregated data on mortality, suicide, and immigration obtained from the Turkish Statistical Institute. The clustering applied on city levels resulted in high levels of bias due to excessive aggregation. In order to reduce the effect of high aggregation, we applied SOM clustering on level one NUTS, or Nomenclature for Territorial Units of Statistics, in which Turkey is divided into 12 regions. We also used U-Matrix for interpreting the results. Comparing the results of our study with the previously published results show that SOM clustering and U-matrix are effective tools for overall interpretation of the determinants and geographical distribution of health related events.

Key Words: Health Informatics; SOM; U-matrix; NUTS

Address for correspondence

Mehrdad A. Mizani,
Informatics Institute, METU
mehrdadmizani@gmail.com