Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 9, n. 1, 2022.

Trabalho apresentado no XLI CNMAC, Unicamp - Campinas - SP, 2022.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics
Preprint

São Paulo City Homicides and Their Surroundings: from Non-Negative Tensor Factorization to Pattern Identification

Jaqueline Silveira[1], Sérgio Adorno[2], Marcelo Nery[3]
NEV, University of São Paulo, São Paulo, SP, Brazil
Germain García[4]
Universidad Católica San Pablo de Arequipa
Afonso Paiva[5], Luis Gustavo Nonato[6]
ICMC, University of São Paulo, São Carlos, SP, Brazil

The type and pattern of crimes change considerably around São Paulo city, presenting regions where crimes occur with great frequency and are typically accompanied by gratuitous violence. Understanding the relationship between homicides and the characteristics of each region has long been a topic of research interest [5]. Several studies have been carried out to understand how socioeconomic variables (population, urban patterns[6], rent values, economic level, and unemployment rate) and urban infrastructure (quality of urban afforestation, slum, tenements, and schools) affect particular crime types [1, 3].

Homicide has been identified as a complex event characterized by many variations, especially in São Paulo city, where the crimes are very different even in near locations. Thus, understanding the urban characteristics around homicides can be of great help once we can identify places in which specific urban characteristics are probably associated with homicides. Therefore, it becomes imperative to understand urban features around homicides in different places to identify the most relevant urban patterns around homicides and the most associated variables.

In this work, in collaboration with a team of sociologists with ample experience in the study of crimes in São Paulo city, we developed a versatile analytical mechanism that relies on **tensor decomposition** [2, 7, 8] to extract urban patterns around homicides from multiple data sources, making the clustering of homicides possible according to those patterns. More specifically, we have gathered socio-economic indicators and information about urban infrastructure in two hundred meters neighborhood of more than ten thousand homicides in São Paulo city. These pieces of information are combined to form an eleven-mode data tensor, which is factorized using a non-negative variant of Tucker decomposition [4]. The decomposition extracts hidden relevant urban patterns, and these patterns are clustered to combine and extract the most representative ones for each group of homicides. Finally, the remaining patterns are appended to the nearest group.

The proposed analytical methodology is more comprehensive than the previous ones, as it makes it possible to handle a large number of variables from different data sources. Our preliminary findings are very encouraging. For example, the extracted urban patterns revealed an association between homicides and afforestation, showing low rates of homicides near wooded places. Our work also revealed an association between homicides and High Urban Patters. It can be explained by the concentration of poor conditions in popular housing, such as the presence of tenements. To validate the obtained results, we have compared our approach against well-established methods.

[1]alva.jaque@usp.br
[2]sadorno@usp.br
[3]mbnery@gmail.com, mbnery@usp.br
[4]germain.garcia@ucsp.edu.pe
[5]apneto@icmc.usp.br
[6]gnonato@icmc.usp.br

 © 2022 SBMAC

More specifically, for each approach, we measure the quality of the clusters by applying clustering evaluation metrics such as Silhouette and Davies Bounding. The results show that our proposal has a better performance than other approaches.

A further contribution of this work is the inclusion of covariates into the model to find patterns and explain the classification of homicides according to those patterns. In fact, our methodology investigates and provides answers to four main questions: i) which are the relations between homicides events and the other variables involved in the analysis? ii) which are the variables that most influence homicides? iii) how to mathematically handle the multiple data sources to uncover patterns? And iv) does tensor decomposition have better performance than other approaches that combine multiple data sources to uncover relevant patterns? As we shall show, the proposed methodology is quite effective in answering the questions above.

Details about the methodology, results, limitations, and future works will also be discussed and shown during the presentation.

References

[1] Joel M. Caplan and Leslie W. Kennedy. **Risk terrain modeling manual: Theoretical framework and technical steps of spatial risk assessment for crime analysis**. Rutgers Center on Public Security, 2010.

[2] Laetitia Gauvin, André Panisson, and Ciro Cattuto. "Detecting the community structure and activity patterns of temporal networks: a non-negative tensor factorization approach". In: **PloS one** 9.1 (2014), e86028.

[3] Andres Gomez-Lievano, HyeJin Youn, and Luis MA Bettencourt. "The statistics of urban scaling and their connection to Zipf's law". In: **PloS one** 7.7 (2012), e40393.

[4] Yong-Deok Kim and Seungjin Choi. "Nonnegative Tucker Decomposition". In: **IEEE CVPR**. 2007, pp. 1–8.

[5] Silas Nogueira de Melo, Lindon Fonseca Matias, and Martin A. Andresen. "Crime concentrations and similarities in spatial crime patterns in a Brazilian context". In: **Applied Geography** 62 (2015), pp. 314–324.

[6] Marcelo Nery, Altay Lino de Souza, and Sérgio Adorno. "Os padrões urbano-demográficos da capital paulista". In: **Estudos Avançados** 33 (Dec. 2019), pp. 5–36. DOI: 10.1590/s0103-4014.2019.3397.002.

[7] Hieu Nguyen and Kiran Garimella. "Understanding International Migration using Tensor Factorization". In: **arXiv preprint:1702.04996** (2017).

[8] S Saritha and G Santhosh Kumar. "An Incremental Boolean Tensor Factorization approach to model Change Patterns of Objects in Images". In: **arXiv preprint:1803.08696** (2018).

© 2022 SBMAC