Spatial subgroup discovery

Ioanna Papagianni

Spatial data mining is the process of exploring possibly unknown and useful patterns from spatial datasets. So far, the focus is on clustering spatial data [6, 2, 8], spatial association rule mining [1], spatial classification [4, 5]. Subgroup discovery is well-studied data mining technique [7, 3] that aims to discover meaningful descriptions of subsets of a dataset among different variables with respect to a target of interest; it aims to find all subgroups within the inductive constraints that reveal a significant deviation in the distribution of the target attribute. To the best of our knowledge we are the first to approach a subgroup discovery task given spatial limitations. Finding subgroups that follow (geo-)spatial thresholds, can lead to more profound discoveries and interpretable explanations.

As a result, the B.Sc. thesis candidate is required to:

- Discover interesting subgroups by exploiting the spatial information from the data,
- apply confusion matrices and different quality measures to the spatial (regression) subgroup findings that summarise the interestingness of the results,
- conduct experiments (random spatial conditions, swap) for statistical validation.

Supervisors: M.Sc. Ioanna Papagianni, dr. Matthijs van Leeuwen

References

[1] R. Bembenik and G. Protaziuk. Mining spatial association rules. In S. T. Wierzchoń M. A. Kłopotek and K. Trojanowski, editors, *Intelligent Information Processing and Web Mining*, pages 3–12, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[2] Dr.E. Chandra. A survey on clustering algorithms for data in spatial database management systems. 2011.

[3] P. González F. Herrera, C. J. Carmona and M. J. del Jesus. An overview on subgroup discovery: Foundations and applications. *Knowledge and Information Systems*, 29:495–525, 12 2011.

[4] P. Ghamisi, M. Dalla Mura, and J. Benediktsson. A survey on spectralspatial classification techniques based on attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 53, 04 2015.

[5] Z. Jiang. A survey on spatial prediction methods. *IEEE Transactions on Knowledge and Data Engineering*, PP:1–1, 08 2018.

[6] E. Kolatch. Clustering algorithms for spatial databases: A survey. 04 2001.

[7] H. M. Proença, T. Bäck P. Grünwald, and M. van Leeuwen. Robust subgroup discovery, 2021.

[8] Z. Shi and L. Pun-Cheng. Spatiotemporal data clustering: A survey of methods. *ISPRS International Journal of Geo-Information*, 8:112, 02 2019.