Special Session Proposal

Geo-Text Data Analysis in Regional Science: recent advances and best practices

Steven Kema, University of Groningen, i.p.s.kema@rug.nl
Harm-Jan Rouwendal, University of Groningen, h.j.rouwendal@rug.nl
Milad Abbasiharofteh, University of Groningen, m.abbasiharofteh@rug.nl
Sierdjan Koster, University of Groningen, sierdjan.koster@rug.nl

Abstract

The exponential growth of 'big data' coupled with enhanced computational capacity and high-performance machine learning techniques provides high-potential research opportunities for the field of Regional Science. User-generated text data is of particular interest, as it often has implicit spatial information embedded within it. These types of geo-text data (Hu 2018) can shed new light on regional economic development, labor market dynamics, and geographies of knowledge production and knowledge relations in a way that traditional secondary data cannot. Some studies in regional studies have already used this new geo-text approach, for example, the digital footprint of interfirm linkages (Kinne and Axenbeck 2020; Abbasiharofteh, Kinne, and Krüger 2021), Twitter data and news items (Ozgun and Broekel 2021), online job vacancies (Acemoglu et al. 2020), but also digitized historical newspaper archives (Peris, Meijers, and van Ham 2021).

Textual data offers numerous advantages. It gives us the ability to create a bottom-up classification of groups and definitions rather than relying on existing definitions. Combined with web scraping techniques, it allows for high-volume flexible, yet robust analysis. There are challenges concerning the reproducibility of results and online-data bias; nevertheless, the geo-text analysis holds great promise for regional science research and should be leveraged much more than currently.

The aim of this special session is to create one of the first interdisciplinary venues for discussing recent advances and best practices in the analysis of geo-text data. We invite conceptual and empirical contributions including but not limited to the following topics:

- developing new regional data using unstructured geo-text data and machine learning techniques,
- mapping and analyzing regional occupation space using job posting data,
- mapping the geographies of knowledge production and knowledge relations using large-scale textual data (e.g., news items and twitter data),

- mining and analyzing web text firm data as an alternative source of data,
- semantic analysis of patents and trademarks using natural language processing techniques,
- investigating historical textual data to map and analyze the rise and decline of innovative places using (e.g., historical newspaper database).

References

Abbasiharofteh, M., J. Kinne, and M. Krüger. 2021. *The Strength of Weak and Strong Ties in Bridging Geographic and Cognitive Distances.* Mannheim: ZEW Discussion Paper No. 21-049.

Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo. 2020. *AI and Jobs*. *Evidence from Online Vacancies.* Cambridge, MA: National Bureau of Economic Research.

Hu, Y. 2018. Geo-text data and data-driven geospatial semantics. *Geography Compass* 12 (11):195.

Kinne, J., and J. Axenbeck. 2020. Web mining for innovation ecosystem mapping. A framework and a large-scale pilot study. *Scientometrics* 28 (4):443.

Ozgun, B., and T. Broekel. 2021. The geography of innovation and technology news - An empirical study of the German news media. *Technological Forecasting and Social Change* 167 (6):120692.

Peris, A., E. Meijers, and M. van Ham. 2021. Information diffusion between Dutch cities. Revisiting Zipf and Pred using a computational social science approach. *Computers, Environment and Urban Systems* 85 (4):101565.