

The Effect of Regional Variation and Resolution on Geosocial Thematic Signatures for Points of Interest

Grant McKenzie, Krzysztof Janowicz, Krzysztof Janowicz

Abstract Computational models of place are a key component of spatial information theory and play an increasing role in research ranging from spatial search to transportation studies. One method to arrive at such models is to extract knowledge from user-generated content e.g., from texts, tags, trajectories, pictures, and so forth. Over the last years, topic modeling techniques such as latent Dirichlet allocation (LDA) have been studied to reveal linguistic patterns that characterize places and their types. Intuitively, people are more likely to describe places such as Yosemite National Park in terms of hiking, nature, and camping than cocktail or dancing. The geo-indicativeness of non-georeferenced text does not only apply to place instances but also place types, e.g., state parks. While different parks will vary greatly with respect to their landscape and thus human descriptions, the distribution of topics common to all parks will differ significantly from other types of places, e.g., night clubs. This aggregation of topics to the type level creates thematic signatures that can be used for place categorization, data cleansing and conflation, semantic search, and so on. To make full use of these signatures, however, requires a better understanding of their intra-type variability as regional differences effect the predictive power of the signatures. Intuitively, the topic composition for place types such as store and office should be less effected by regional differences than the topic composition for types such as monument and mountain. In this work, we approach this regional variability hypothesis by attempting to prove that all place types are aspatial with respect to their thematic signatures. We reject this hypothesis by comparing the signature similarities of 316 place types between major cities in the U.S. We then select the most and least varying place types and compare them to thematic signatures from regions outside of the U.S. Finally, we explore the effects of LDA topic resolution on differences between and within place types.

Keywords Semantic signature, Point of interest, Place type, Geosocial, Topic modeling