Event Geography Modeling

By May Yuan

Geospatial research has long focused on continuants (e.g. themes and features) but has paid little attention to occurrents (e.g. events and activities). In contrast, occurrents are central to quantitative social sciences, such as public health, political science, criminology, and sociology. Box-Steffensmeier and Jones (2004) summarized the principles of event history analysis and longitudinal methods predict the likelihood of the occurrence of a specific event (a.k.a. time-to-event variable or survival variable) to individuals, which can be a person, a community, or a nation. While "spatial turns" are prevailing in social sciences, the "turns" tend to apply geospatial methods for continuants directly to occurrents. For example, hot spot analysis is popular in identifying clusters of disease infections, crime events or traffic accidents without considerations of event characteristics, such as time of occurrences and cascading effects.

In this study, we explore the idea of event geography modeling as the first step to compliment the well-established event history modeling. Event history modeling takes account not only whether the event of interest will occur but when. As such, event geography modeling shall address where the event of interest will occur subject to the assumption that the specific event does not occur everywhere and its occurrence is not random. An apparent example is that traffic accidents only take place on transportation networks, so hot spot analysis should consider discrete space on a network, not a continuous Cartesian space and time. While event history modeling considers individuals as persons or communities, event geography modeling considers individuals as locations or areas. Certain locations are at a higher risk than others of experiencing the event of interest, and the risk of experiencing an event at a certain time can be predicted with a set of covariates.

Hot spot analysis or generally spatial clustering analysis conceptually focus on the distribution and prediction of points in space and time with event points as the unit of analysis. Alternatively, the proposed event geography modeling uses space-time atom as the unit of analysis. Goodchild et al. (2007) outlined a general representation theory for geographic dynamics, in which space-time atoms represent the smallest unit that can form objects or fields through strategic aggregation of attributes and/or relationships in space and time, Our proposed event geography modeling subcribes to the theory but proceeds in a reversed fashion: we start with a space-time aggregate of event occurrences on a network, with which we need to define proper space-time atoms for analysis. A space-time atom for network-based likelihood modeling of event occurrences is network's smallest unit, which contains a maximum of one event point. By focusing on space-time atoms of one or no event, we transform the problem of modeling spatial point processes to estimating the chance of one event among binary space-time atoms. Hereafter, we refer the binary space-time atoms as event space-time atoms and nonevent space-time atoms, respectively. Spatial analyses of the event distribution on a network can then base on continuity of event space-time atoms and connectivity of event and nonevent space-time atoms.

Event geography modeling is based on the premise that site and situtation characteristics of "where" help explain why and how events occur. Ullman (1954) coined the idea of "geography as spatial interaction" and emphasized the important interplay between site and situation. Site is about the location of an area as characterized by its local, underlying physical environments. Situation refers to how an area's location relates to others' locations, the effects of one area to another, and the connection between areas. Our proposed conceptual framework makes two extensions to Ullman's thoughts in site and situation. First, we assume that site characteristics reflect geographic assets that afford certain events and hence signal where the events are more likelily to occur than other locations. The conceptual framework consititutes the following procedures in event geography modeling:

1. Determine the space-time atom of events on a discrete space (e.g. a network) in such a way that each space-time atom has one or zero event. This is the binary dependent variable for likelihood modeling.
2. Determine the site and situation that characterize every space-time atom and its spatial interactions with other space-time atoms. These are the independent variables for the likelihood modeling.
3. Develop a model to predict the likelihood of having an event on every space-time atom on the network. Since the dependent variable is binary, a logistic regression model is applicable to predict the probability of an event space-time atom across the network. Other models are also possible, such as Bayesian binary regression model, classification and regression trees or random forest algorithm.

The proposed approach is generic and applicable to point events on continuous or discrete space. Using data from Dallas, Texas, USA, we take the proposed event geography modeling approach to estimate the likelihood of traffic accidents based on binary (event or non-event) space-time atoms of 100-m road segments and 1-hour intervals. We choose 12 variables representing time, site characteristics, and situational conditions to develop logistic regression and random forest models. The traffic accident data on even days were used for model construction and data on odd days for model testing. Both models result in comparable accuracy at 84.11% (logistic regression) and 85.42% (random forest) with significant differences in the spatial patterns of how site and situation relate to traffic accidents. The difference signals the dynamic influence of site and situation characteristics on the event likelihood over time.

The proposed event geography modeling uncovered the dynamic influence of site and situation characteristics beyond what existing geospatial analysis can distill. Time plays an interesting role in both models on the traffic risk prediction. The interaction between time and major intersections in the logistic regression model reflects that making a turn at a major intersection is always dangerous, but the danger increases with higher traffic flow. The interaction between time and space syntax is complex. In the logistic regression model time merely amplifies the positive relationship between time and both integration and choice. That time alone is insignificant adds credence to the idea that traffic accident patterns reflect the larger pattern of life in the city. In the random forest model, the effect of both choice and integration on accident likelihood may completely reverse itself depending on the time of day. At busy times of day, high integration and high choice increase the likelihood of accidents, but at night high values of both variables decrease the likelihood of accidents. Road segments of high integration and high choice are most associated with major roads with four or more lanes. During the overnight hours, less traffic on wider roads leads to the lower likelihood of traffic accidents compared to other road segments. During the daytime, drivers are often in the middle of their journeys, in central parts of the city and on main thoroughfares, but at the start and end of the day, they are closer to home, which in Dallas is typically in the suburbs. The spatial risk distributions predicted by the random forest model follow such a shift in the distribution of high-risk road segments.

The case study showed the usefulness of the proposed event geography modeling in predicting space-time risk of and explanatory variables to traffic accidents. Future research is planned to test the conceptual framework on other event types and refine the framework to consider space-time risk of moving events on a network, such as parades, protests, and marathons.