

Detecting location spoofing in social media:

A Bayesian time geography approach

Daniel Sui and Bo Zhao

Extended Abstract. Location spoofing refers to the emerging geographic practices that allow users to hide their true geographic locations. The proliferation of location spoofing in recent years has stirred debate about the reliability and convenience of crowd-sourced geographic information and the use of location spoofing as an effective countermeasure to protect individual geo-privacy and national security. However, these polarized views do not contribute to a solid understanding of the complexities of this trend. As today, we still lack a robust method for detecting location spoofing and a holistic understanding about its multifaceted implications. The primary goal of this paper is to develop a quantitative approach for detecting location spoofing.

The algorithm to detect fake locational information l_s is based on the filter-refine strategy. In general, this strategy consists of two steps. First, a set of candidates is filtered from the original data. This new set of candidates is a superset of the answer set, which consists of both the actual answer and the false ones; secondly, each candidate from the filter step is then analysed according to some other constraints in order to produce the answer set by eliminating the false ones. This strategy is employed as shown in Figure 1. Firstly, the detecting algorithm singles out a set of space-time (ST) path candidates that violate human activity constraints through ST cones. Then, for each ST path candidate, a Bayesian time geographic approach is conducted to calculate the probabilistic distribution of visiting different places in the research area for the author of the path. After comparing the probabilistic values of the two pieces of locational information (which link to the ST path), we deem the one with a lower visiting probability is more likely to be the fake locational information. In the following paragraphs, the flowchart is explained in greater detail.

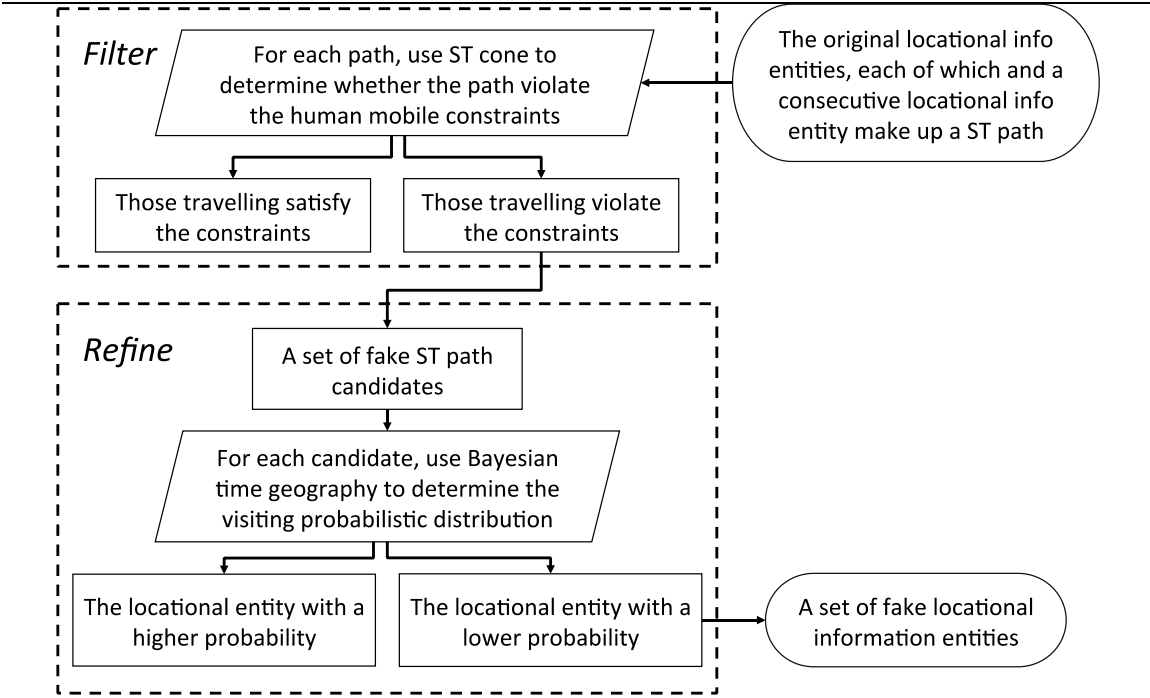


Figure 1.

We tested our approach using millions of geo-tagged tweets, which is one of the major sources of big data related to human activities. The results indicate that the proposed approach can successfully detect certain types of location spoofing. Based on the empirical results, this study further examines the possible motivations of location spoofing and its inherent uncertainties. Rather than simply neglecting the spoofing phenomenon, this paper calls for directly confronting this thorny issue, especially when any arguments or policies are drawn from geospatial big data. Only then can we promote more effective and trustworthy uses of big data.

Keywords: location spoofing, data quality, big data, uncertainty, social media