

Cluster Analysis on Tweet-Generated Hurricane Evacuation Trajectories

Chenxiao Guo

In this project I will use the geo-tagged Tweet data related to this specific hurricane season to generate the trajectories of users, and then conduct a cluster analysis to reveal the spatiotemporal pattern of evacuation behaviors, which is helpful to better understand the evacuation pattern and further provide guide in disaster management.

I generated the trajectories from some Twitter users during the interested hurricane periods. After some necessary data processing, I get 17704 valid trajectories around the United States. Then I generate the cluster analysis to figure out how the spatiotemporal clusters match the real condition of hurricane events. The variable I consider are the projected coordinates of origin and destination, as well as the time of them, and then I normalized them to the scale of 0-1.

Before conducting the K-Means cluster analysis, the chart of Between-Group Sum-of-Square shows that 8 could be a good value for K. The result of the analysis includes a cluster whose spatial and temporal characteristics match the condition of Hurricane Harvey, and another one matches Hurricane Irma.

Result indicates that the Cluster around Harvey has broader directions for destination while the Cluster around Irma are more concentrated to the north and west, especially towards northwest. Also, from the Cluster around Irma we notice that a lot of people are traveling at a medium distance (typically towards northwest) and there is no such pattern in the Cluster around Harvey. The overall amount (size) also shows a significant difference. As for the temporal pattern, the Cluster around Irma has a larger average time range (origin and destination) as well as a longer average time duration, compared to the Cluster around Harvey.

In this study, the prior knowledge plays a key part in terms of both time and space, which helps to find the matching clusters. However, this way only works for very general cluster (large size) but cannot work for detecting potential cluster from the hurricane events or detecting the possible heavy transportation flow to help manage the traffic jam. In the future I could explore other methods to detect cluster in a smaller spatial scale.