Statistical Significance of Discriminative Sub-trajectory

Vo Nguyen Le Duy  
Takuto Sakuma  
Taiju Ishiyama  
Hiroki Toda  
Kazuya Arai  
Masayuki Karasuyama  

Nagoya Institute of Technology

Yuta Okubo  
Masayuki Sunaga  

Sompo Japan Nipponkoa Insurance Inc

Yasuo Tabei  
Ichiro Takeuchi  

RIKEN Center for Advanced Intelligence Project (RIKEN AIP)  
Nagoya Institute of Technology, RIKEN AIP

Abstract

We study the problem of discriminative sub-trajectory mining. Given two groups of trajectories, the goal of this problem is to extract moving patterns in the form of sub-trajectories which are more similar to sub-trajectories of one group and less similar to those of the other. We propose a new method, called Statistically Discriminative Sub-trajectory Mining (Stat-DSM), to evaluate the statistical significance of the findings in the discriminative sub-trajectory mining problem. An advantage of the Stat-DSM method is that the statistical significance of the extracted sub-trajectories are properly controlled in the sense that the probability of finding a false positive sub-trajectory is smaller than a specified significance threshold alpha (e.g., 0.05), which is indispensable when the method is used in scientific or social studies under noisy environment. Finding such statistically discriminative sub-trajectories from massive trajectory dataset is both computationally and statistically challenging. In the Stat-DSM method, we resolve the difficulties by introducing a tree representation among sub-trajectories and running an efficient permutation-based statistical inference method on the tree. To the best of our knowledge, Stat-DSM is the first method that can efficiently extract statistically discriminative sub-trajectories from massive trajectory dataset.

Keywords: trajectory mining, significant pattern mining, statistical testing, multiple testing