Influence maximization problem to find the influential location

Sun Yehong

Shanghaijiaotong University

syhroy7@gmail.com

May 27, 2018

Overview

1. Introduction
   - Influence maximization in social network
   - Model the Influence maximization problem in real place
2. Design And Implementation
   - Data process
   - Experiment
   - Result and analysis
3. Future work and conclusion
   - Future work and conclusion
   - References

Influence maximization problem

Information can propagate from one node to another node through a link on a social network,which makes it an important research issue to find influential nodes for the spread of information through a network represented by a directed graph. These combinatorial optimization problem was called the influence maximization problem.

IC model

Definition

In this model, we must specify a real value p(u,v) $\in$ [0, 1] for each directed link (u, v) in advance. Here, p(u,v) is referred to as the propagation probability through link (u, v).

Diffusion process

Given an initial set A of active nodes, when node u first becomes active at step t, it has a chance to activate each currently inactive child v with probability p(u,v). If succeeds, then v will become active at step t+1.If multiple parents of v first become active at step t, then their activation attempts are sequenced in an arbitrary order, but performed at step t. Whether or not u succeeds, it cannot make any further attempts to activate v in subsequent rounds. The process terminates if no more activations are possible.

Finding the influential locations

- Create the IC model using the dataset of taxi trajetories in Shanghai.
- Cluster the start points and end points to get points set V.
- Using the relationship of start points and end points to get the set of edges E.
- Analysis the IC model we get to find the influential locations

Pre-process

- trajectory mapping, I maps the raw trajectories onto the corresponding spatial network
- spatial network construction, The initial points set has too many points which are not representative then I clustered the locations based on the trajectories, and then constructs the spatial network.
- Graph partitioning is used to divide a graph into several chunks while satisfying certain constraints and objectives

Result

Figure: The mapping of the trajectories

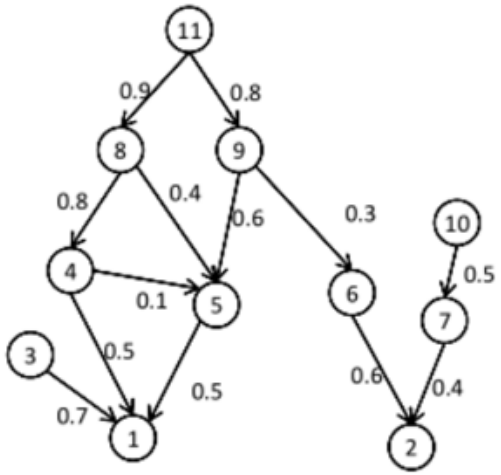

Figure: A simple IC model

- After my pre-process, the points set has been reduced to 274 points.
- In the experiment, I compared the two algorithm to verify the accuracy of my opinion. The first algorithm is the number of occurrences first. Another algorithm is the greedy hill-climbing algorithm.
- I test these algorithm with the initial graph and the initial number of points k. And we choose different k 1,5,10,50,100.

Result and analysis

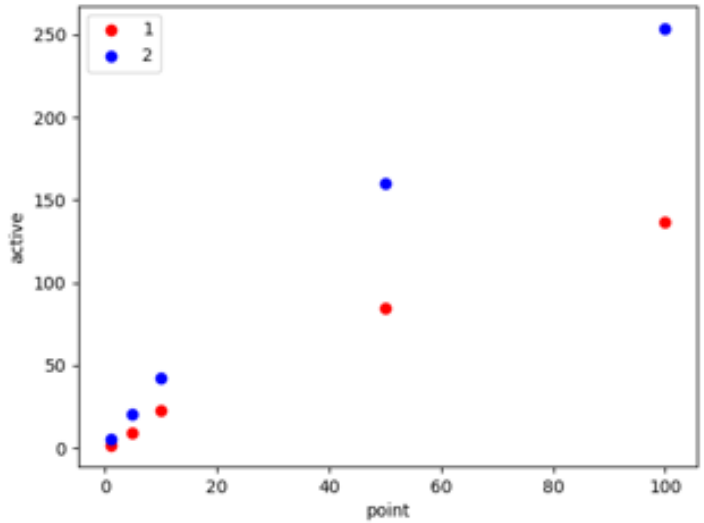

Figure: Result of the experiment

We can see that the points activated in algorithm 2 is much lager than those in algorithm 1.

Conclusion and future work

- We can see that my algorithm is much better than the first algorithm. In the real life, this will be useful for the resource allocation and the chain store layout. So in my opinion, this work is useful and the result is pretty good.
- There are many new model in the influence maximization problem, the complexity of some model algorithm even can be O(m) called Degree Heuristic.
- I want to improve the clustering algorithm and add the influence of the POI

References

Masahiro Kimura

Extracting Influential Nodes for Information Diffusion on a Social Network

www.aaai.org

Yuhong Li,Jie Bao,Yanhua Li,Yingcai Wu,Zhiguo Gong, Yu Zheng

Mining the Most Influential k-Location Set From Massive Trajectories

http://dx.doi.org/10.1145/2996913.2997009

Yu Zheng, Lizhu Zhang, Xing Xie

MaMining Interesting Locations and Travel Sequences from GPS Trajectories

Tong Xu

Taxi Driving Behavior Analysis in Latent Vehicle-to-Vehicle Networks: A Social Influence Perspective

http://dx.doi.org/10.1145/2939672.2939799

Thanks for listening!