Reverse Nearest Neighbor Heat Maps: A Tool for Influence Exploration

Yu Sun †1, Rui Zhang †2, Andy Yuan Xue †3, Jianzhong Qi †4, Xiaoyong Du ‡5

† The University of Melbourne
{1 sun.y, 2 rui.zhang, 3 andy.xue, 4 jianzhong.qi}@unimelb.edu.au

‡ Renmin University of China and Key Lab of DEKE, MOE, China
5 duyong@ruc.edu.cn

May 19th 2016
Outline

1. Motivation and Problem Definition
2. Algorithms
3. Experiments
4. Conclusion and Future Work
Motivation and Problem Definition

1. Motivation
   - Problem Definition

2. Algorithms
   - A Baseline Algorithm
   - The CREST Algorithm
   - Analysis
   - CREST In Other Settings

3. Experiments
   - Set-up
   - Results

4. Conclusion and Future Work
Decision Making for Locations

Where to put the ATM?

Where to go next?

Y. Sun, R. Zhang, A. Y. Xue, J. Qi, and X. Du

RNN Heat Maps: A Tool for Influence Exploration
Quantitative and Qualitative Factors

Decision Making

- Many quantitative and qualitative factors

An Important Quantitative Factor

- Distance to users and facilities, e.g., ATMs or taxis
- Model this factor: the reverse nearest neighbors
Quantitative and Qualitative Factors

Existing Algorithms
- Find locations having the largest RNN set

Implicit Assumptions
- Assume this quantitative factor is the only factor in decision making
- Assume the quantity is only measured by the size of the RNN set

Be Careful
- These assumptions are generally not true in many applications
Location Exploration

Quantitative Factor
- any real-valued function on the RNN set
- e.g., consider the demographic information or social ties

Qualitative Factors
- cannot be effectively quantified
- subject to decision maker’s judgments
- e.g., area safety or convenience of transportation
Location Exploration

Heat Map of New York City

Satellite Map of New York City
Outline

1 Motivation and Problem Definition
   - Motivation
   - Problem Definition

2 Algorithms
   - A Baseline Algorithm
   - The CREST Algorithm
   - Analysis
   - CREST In Other Settings

3 Experiments
   - Set-up
   - Results

4 Conclusion and Future Work

Y. Sun, R. Zhang, A. Y. Xue, J. Qi, and X. Du

RNN Heat Maps: A Tool for Influence Exploration
RNN Heat Map Problem

Definition (RNNHM Problem)

Given two sets of points $\mathcal{O}$ and $\mathcal{F}$, a distance metric, and an influence measure which is a real-valued function on the RNN set, find for each point its influence value, i.e., the heat value.
A Simple Superimposition Cannot Work

RNN Heat Maps: A Tool for Influence Exploration

Y. Sun, R. Zhang, A. Y. Xue, J. Qi, and X. Du
Necessity of Systematically Building A Heat Map

Data points

Superimposition

Heat map with size measure

Heat map with other measure

Y. Sun, R. Zhang, A. Y. Xue, J. Qi, and X. Du

RNN Heat Maps: A Tool for Influence Exploration
Outline

1 Motivation and Problem Definition
   - Motivation
   - Problem Definition

2 Algorithms
   - A Baseline Algorithm
   - The CREST Algorithm
   - Analysis
   - CREST In Other Settings

3 Experiments
   - Set-up
   - Results

4 Conclusion and Future Work
Arrangement View

Arrangement of NN-circles

- Face $\Rightarrow$ region
- Points in a region have the same RNN set
- Heat map $\Leftarrow$ compute RNN set for each region
Baseline: Side Extension

A Straightforward Approach

- Extend the sides of each NN-circle
- Sort, and locate each subdivided region
- Use point enclosure query to obtain RNN set

Problem: Too slow!
Outline

1 Motivation and Problem Definition
   - Motivation
   - Problem Definition

2 Algorithms
   - A Baseline Algorithm
   - The CREST Algorithm
   - Analysis
   - CREST In Other Settings

3 Experiments
   - Set-up
   - Results

4 Conclusion and Future Work
Avoiding Point Enclosure Queries

Idea Is Quite Straightforward

- obtain the RNN set from adjacent regions
- such operation can be easily supported by a plane sweep
Reducing the Times of RNN Computation

Locating Change Intervals

- repeated computation for many regions
- locate the area where the RNN set is changed
- change of NN-circles, which are either inserted or removed from the sweep line
Reducing the Times of RNN Computation

Caching Base Sets
- need a base set for each change interval
- caching the RNN sets of previous events
- maintained and associated with the elements in the sweep line

Correctness
- prove by induction that the algorithm is correct
Outline

1 Motivation and Problem Definition
   - Motivation
   - Problem Definition

2 Algorithms
   - A Baseline Algorithm
   - The CREST Algorithm
   - Analysis
   - CREST In Other Settings

3 Experiments
   - Set-up
   - Results

4 Conclusion and Future Work
Complexity of CREST

**Complexity**

The CREST algorithm solves the problem in $O(n \log n + r \lambda)$ time with $O(n \lambda)$ space.

**Lower Bound**

$\Omega(n \log n + r \lambda^*)$ is a lower bound of the problem.
Outline

1 Motivation and Problem Definition
   - Motivation
   - Problem Definition

2 Algorithms
   - A Baseline Algorithm
   - The CREST Algorithm
   - Analysis
   - CREST In Other Settings

3 Experiments
   - Set-up
   - Results

4 Conclusion and Future Work
Monochromatic RNNs
- only have one set of points
- a special case of bichromatic RNNs

RNNHM with L1 Distance
- rotate the coordinate system by $\pi/4$
- diamonds become squares
RNNHM with L2 Distance

Modifications

- Events: extreme, center, and intersection points
- Maintain a proper vertical order of arcs
- Update the coordinates of these arcs for each event
Outline

1 Motivation and Problem Definition
   - Motivation
   - Problem Definition

2 Algorithms
   - A Baseline Algorithm
   - The CREST Algorithm
   - Analysis
   - CREST In Other Settings

3 Experiments
   - Set-up
   - Results

4 Conclusion and Future Work
Data Sets and Distances

Experiment Set-up

- Points of interests in New York City and Los Angeles from Foursquare
- Experiment with L1 and L2 distances
- Implement the algorithm with C++ and use a desktop with a 3.4GHz Intel i7-2600 CPU

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYC</td>
<td>128,547</td>
<td>points-of-interest in New York City</td>
</tr>
<tr>
<td>LA</td>
<td>116,596</td>
<td>points-of-interest in Los Angeles</td>
</tr>
</tbody>
</table>
Outline

1. Motivation and Problem Definition
   - Motivation
   - Problem Definition

2. Algorithms
   - A Baseline Algorithm
   - The CREST Algorithm
   - Analysis
   - CREST In Other Settings

3. Experiments
   - Set-up
   - Results

4. Conclusion and Future Work
Showcasing Real-World Heat Maps

Heat Map of Los Angeles


Satellite Map of Los Angeles
Performance of CREST with L1 Distance

(a) LA

Cardinality of O

(b) NYC

Cardinality of O

(c) LA

CPU time (ms)

(d) NYC

Ratio |O|/|F|

Y. Sun, R. Zhang, A. Y. Xue, J. Qi, and X. Du

RNN Heat Maps: A Tool for Influence Exploration
Performance of CREST with L2 Distance

(e) LA

(f) NYC

(g) LA

(h) NYC

CPU time (ms) vs. Ratio $|O|/|F|$

CPU time (ms) vs. Cardinality of $O$

Pruning CREST-L2
Take Home Message One

- The heat map relaxes several assumptions of existing methods for support decision making of locations, and can effectively assist influence exploration.

Take Home Message Two

- The proposed CREST algorithm can efficiently generate such heat maps in various settings.