Top-k Most Incremental Location Selection with Capacity Constraint

Yu Sun
Renmin University of China
Outline

- Motivation and Introduction
- Related Work
- Definitions
- Algorithms
- Experiment Results
- Conclusion
Outline

• Motivation and Introduction
• Related Work
• Definitions
• Algorithms
• Experiment Results
• Conclusion
Motivation and Introduction

• **Question**: location selection for facilities that bear capacity constraint

• A facility set of $W$, a customer set of $R$. $W$ provides service for $R$.

• When the number of customers increases, more facilities are needed.

• $P$ is a set of candidate locations, then how to get the most promising candidates?
Motivation and Introduction

Applications:
- Online shopping
- Mobile Service
- ATM
- …
Motivation and Introduction

- $C(w1) = 5$, $C(w2) = 2$, $C(w3) = 4$
Motivation and Introduction

- After p1 is added, here $C(p1) = 4$
Motivation and Introduction

- **Challenging:**
  - (1) New influence set
    - not only related to *customer locations*, but also related to *locations of facility*
  - (2) Dealing a large amount of input data
    - NE contains **123,593** points of postal addresses
    - We generate $|R|=200,000, |W|=20,000$ and $|P|=10,000$
Outline

- Motivation and Introduction
- Related Work
- Definitions
- Algorithms
- Experiment Results
- Conclusion
Related Work

- Korn et. al. *Influence sets based on reverse nearest neighbor queries*. *SIGMOD*’00.
- Xia et. al. *On computing top-t most influential spatial sites*. *VLDB*’05.
- However, they **fail to** take capacity constraint into consideration.
- Though capacity is considered, it is designed for profile-matching applications and **not applicable**
Outline

• Motivation and Introduction
• Related Work
• Definitions
• Algorithms
• Experiment Results
• Conclusion
Definitions

- The **service quality** is the total number of customers who are provided reliable service.
- The **candidate’s increment** equals new service quality – old service quality.
Definitions -- Basic Solution

• The definition gives a basic solution
• Suppose we can calculate the service quality
• Scan the candidate set, add each candidate to the facility set
• Calculate the new service quality
• Pick out candidates with top k increments
• BUT IT IS TOO SLOW!
Outline

• Motivation and Introduction
• Related Work
• Definitions
• Algorithms
• Experiment Results
• Conclusion
Algorithms

• Our proposed algorithms improve the basic solution from the following aspects:

• (1) Redefine the Influence Set, which contains only those affected customers and facilities, to effectively prune the search space

• (2) Make use of Nearest Facility Circles and spatial indices to efficiently get influence sets
Algorithms

• example of an influence set
Algorithms

- example of a nearest facility circle (NFC)

- NFC helps to get influence sets without NN queries
Algorithms

• example of finding influence set with NFC
Algorithms

- The time complexity of the basic solution is $O(n^3)$
- The influence sets reduce the complexity to $O(n^2)$
- The NFCs further decrease the complexity to $O(n\log n)$
- We call them basic, pruning and index respectively.
Outline

• Motivation and Introduction
• Related Work
• Definitions
• Algorithms
• Experiment Results
• Conclusion
Experiment results

• Comparison

- Real Dataset
- Synthetic Dataset
Experiment results

- Scalability
Outline

• Motivation and Introduction
• Related Work
• Definitions
• Algorithms
• Experiment Results
• Conclusion
Conclusion

- Formulate the top-k most incremental location selection query
- Propose pruning techniques and an efficient algorithm to answer the query
- Experiment results confirm the effectiveness of the algorithm
Q&A

• Questions?

• Thank you for your time!