Optimal Mobile Users for Long-term Environmental Monitoring by Crowdsourcing

Juan Li, Jie Wu, and Yanmin Zhu

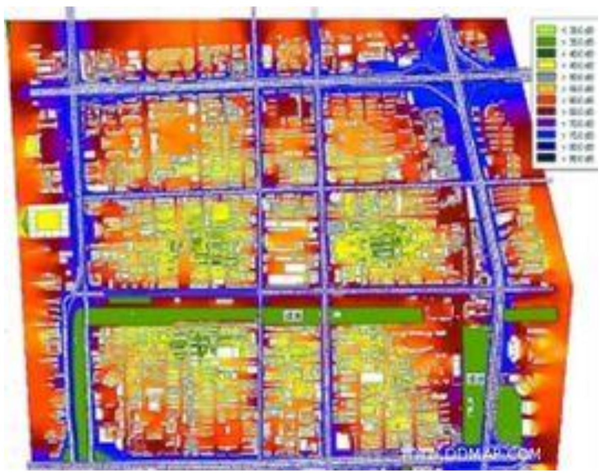
Shanghai Jiao Tong University

Temple University

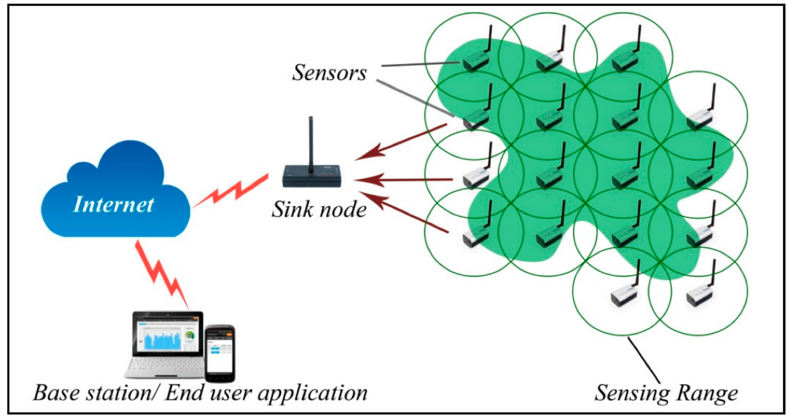
Road Map

- Introduction
- Model and Formulation
- Observation and Idea
- Algorithms
- Experiments
- Conclusion

Background

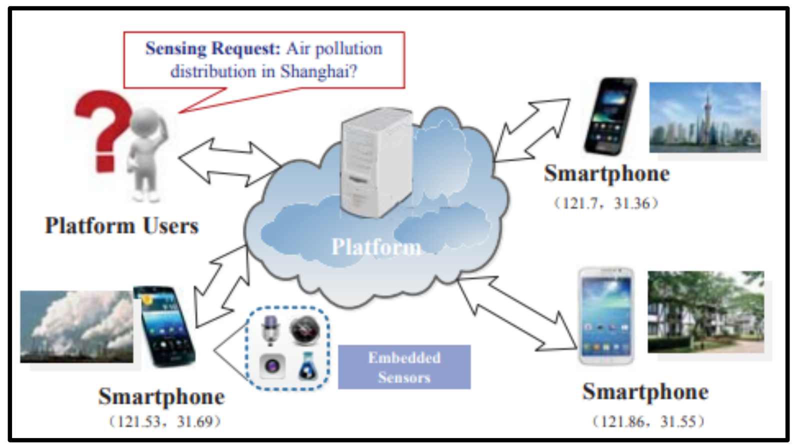


Air pollution map



Pre-deployed sensor network

- High deployment and maintenance cost
- Monitoring is coarse-grained

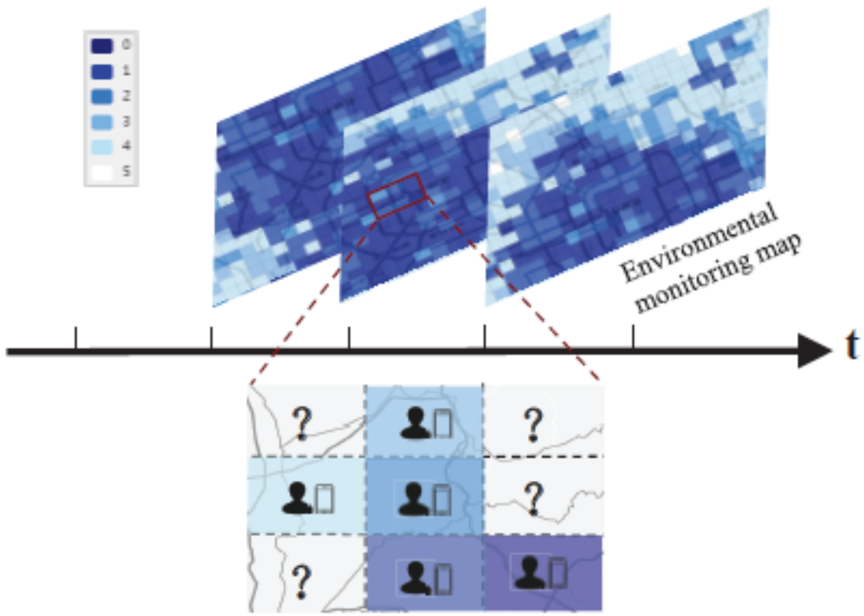


Crowdsourcing over mobile devices

- Cheap;
- Fine-grained monitoring

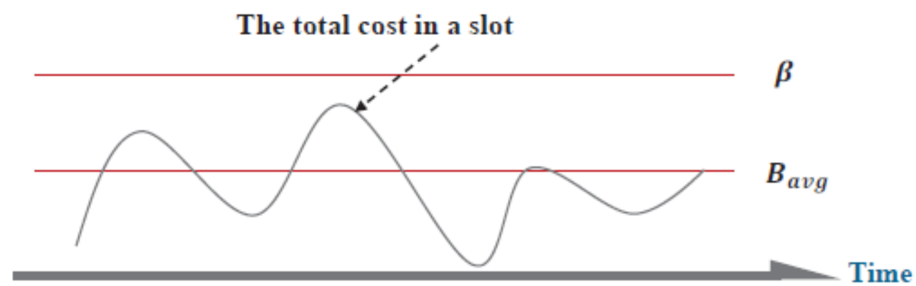
Problem

The air pollution map is updated in each time slot



Some grids do not have measurements because of the limited budget.

How to get accurate monitoring maps under the limited budget?

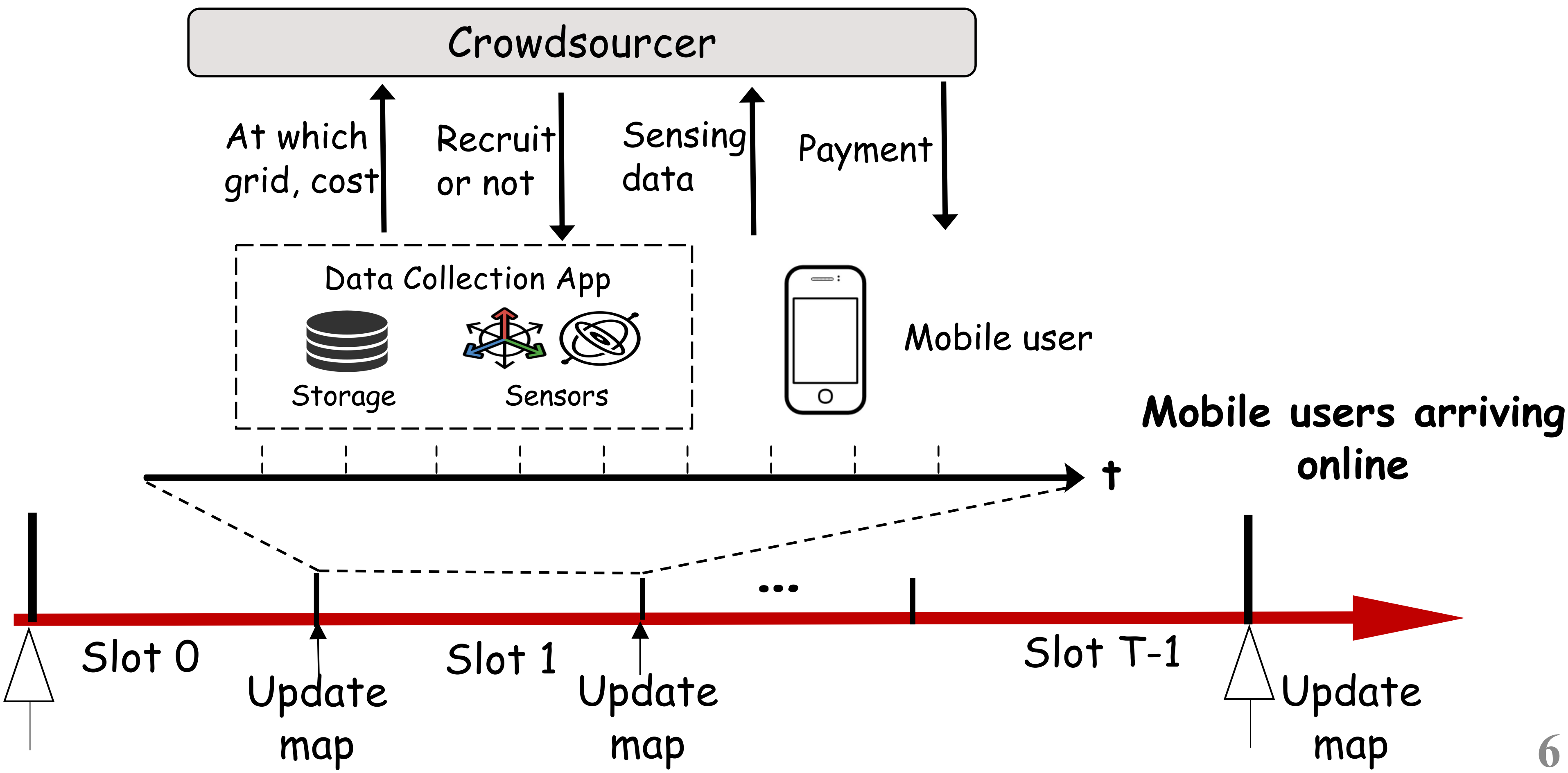


- The average payment should be below B_{avg}.
- The payment in each slot cannot exceed β.

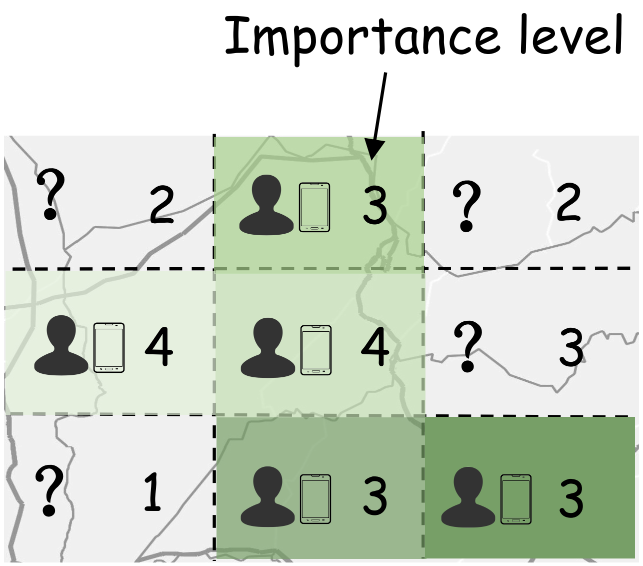
Road Map

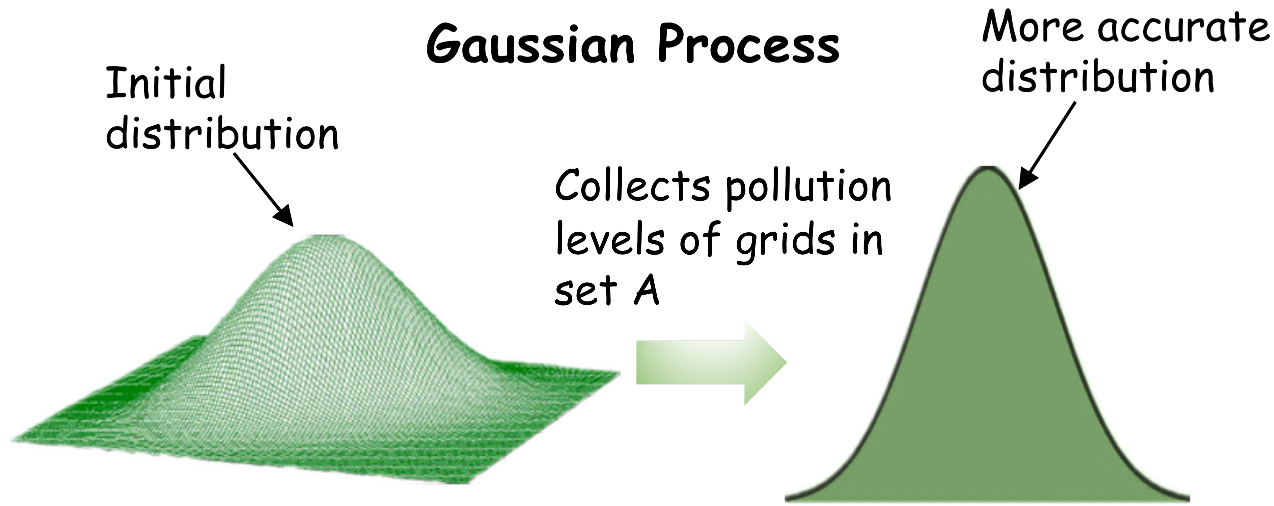
- Introduction
- Model and Formulation
- Observation and Idea
- Algorithms
- Experiments
- Conclusion

Model and Formulation



Objective in a slot -Data Utility





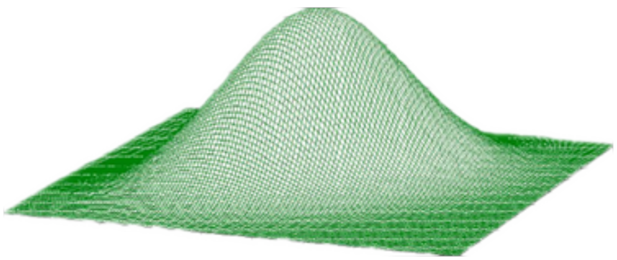
- Measurements at important grids can be collected.
- The crowdsourcer is sure of inferences of other grids.

Air pollution levels of all girds in the universal set U~$N(\vec{\mu}, \Sigma)$(known)

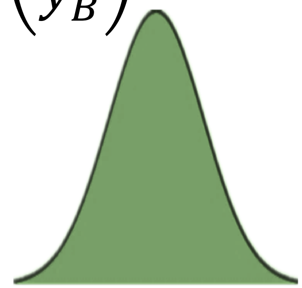
Air pollution levels of girds in set B=U\A ~$N(\vec{\mu}_A, \Sigma_A)$

Gaussian Process

Air pollution levels of girds in the universal set U: $\vec{y} = \begin{pmatrix} \overrightarrow{y_A} \\ \overrightarrow{y_B} \end{pmatrix}$



$\overrightarrow{y_A}$ is known



$\begin{pmatrix} \overrightarrow{y_A} \\ \overrightarrow{y_B} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \overrightarrow{u_A} \\ \overrightarrow{u_B} \end{pmatrix}, \begin{pmatrix} \Sigma_{AA}, \Sigma_{AB} \\ \Sigma_{BA}, \Sigma_{BB} \end{pmatrix}\right)$ (known)

Initial entropy of $\overrightarrow{y_B}$:

$$\mathrm{H}(\overrightarrow{y_B}) = \frac{1}{2} \ln[(2\pi e)^{|B|} |\Sigma_{BB}|]$$

$$\overrightarrow{u_{B|A}} = \overrightarrow{u_B} + \Sigma_{BA} \Sigma_{AA}^{-1} (\overrightarrow{y_A} - \overrightarrow{u_A})$$

$$\Sigma_{B|A} = \Sigma_{BB} - \Sigma_{BA} \Sigma_{A}^{-1} \Sigma_{AB}$$

Conditional entropy of $\overrightarrow{y_B} | \overrightarrow{y_A}$:

$$\mathrm{H}(\overrightarrow{y_B} | \overrightarrow{y_A}) = \frac{1}{2} \ln[(2\pi e)^{|B|} |\Sigma_{B|A}|]$$

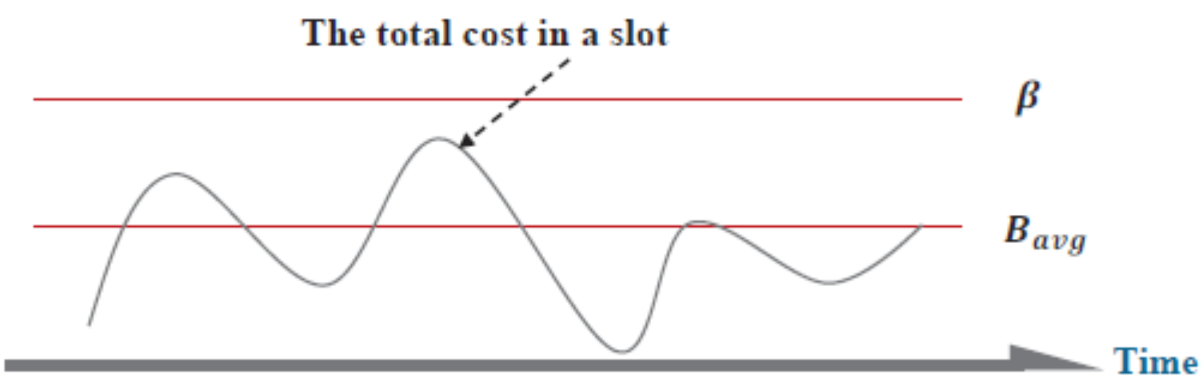
Data utility in a slot= entropy decrease of unknown grids + sum of importance levels of known grids

Problem Formulation

Max $\frac{1}{T}\lim_{T->\infty} \Sigma_1^T$(data utility in slot t)

S.t. The total payment in each slot $\leq \beta$

The average of the total payment in each slot $\leq B_{avg}$



Two challenges:

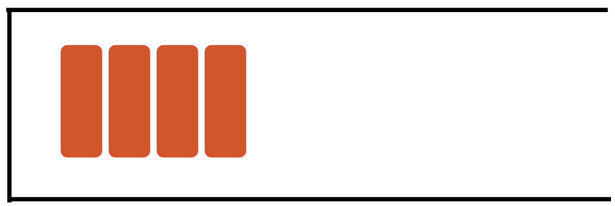
- Online problem
- NP-hard problem

Road Map

- Introduction
- Model and Formulation
- Algorithm Design
- Experiments
- Conclusion

Decoupling the Long-term Problem

Lyapunov optimization



The virtual queue length Q(t) represents the over used budget in the past

- Q(t+1)=Q(t) +max[the total payment in slot t-B_{avg},0]
- $\Delta Q(t) = E\left\{\frac{1}{2}Q^2(t) - \frac{1}{2}Q^2(t-1)\middle|Q(t-1)\right\}$ measures the increase of the queue length between two slots.

Decoupling the Long-term Problem

The long-term problem

Max the average data utility
S.t. The upper bound of the payment in each slot
The average budget constraint

The short-term problem in slot t

Max G_t=V *data utility in slot t- $\Delta Q(t)$
S.t. The upper bound of the payment in slot t is β

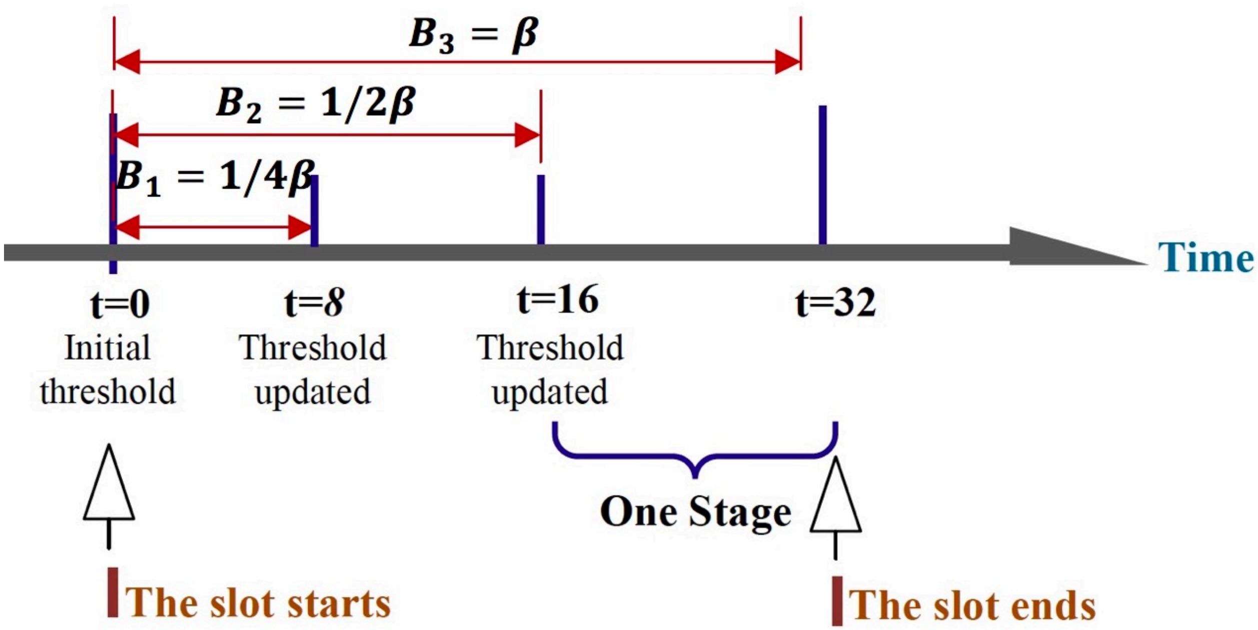
Theoretical performance:

- The average budget constraint can be satisfied as long as the time is long enough.
- If $G_t \geq eG_t^*$(e is the competitive ratio for the one-slot problem),

$$\frac{1}{T}\lim_{T->\infty} \Sigma_1^T \; (data\ utility\ in\ slot\ t)$$
$$\geq e\times the\ optimal\ average\ utility - D/V$$

a small constant

Online Mobile User Selection Algorithm in a slot



- When a mobile user comes, the crowdsourcer recruits the user if
  - The marginal contribution/the cost ≥ a threshold
  - the remaining budget can afford the cost

- The threshold is updated at the end of each stage i.
  - Using the users arriving before and B_i as input, we can obtain an offline objective G_i by a greedy strategy (greedy metric is contribution/cost of user)
  - The threshold is updated as $G_i/(6B_i)$

Road Map

- Introduction
- Model and Formulation
- Algorithm Design
- Experiments
- Conclusion

Experiments

- **Data sets:** the air pollution data in Beijing; the real human trajectory data.
- **Baselines**: Cost First, Shortsighted UPR and Shortsighted AVG
- **Settings:**
  - the number of slots is 2800;
  - the weight V is 10;
  - the upper bound of the one-slot payment is $700;
  - the average budget is $500;
  - the area of one grid is 5km*5km;
  - the importance levels are {1,2,3,4,5};
  - the cost of each mobile user is form $0.2 to $1.5.

Experiments

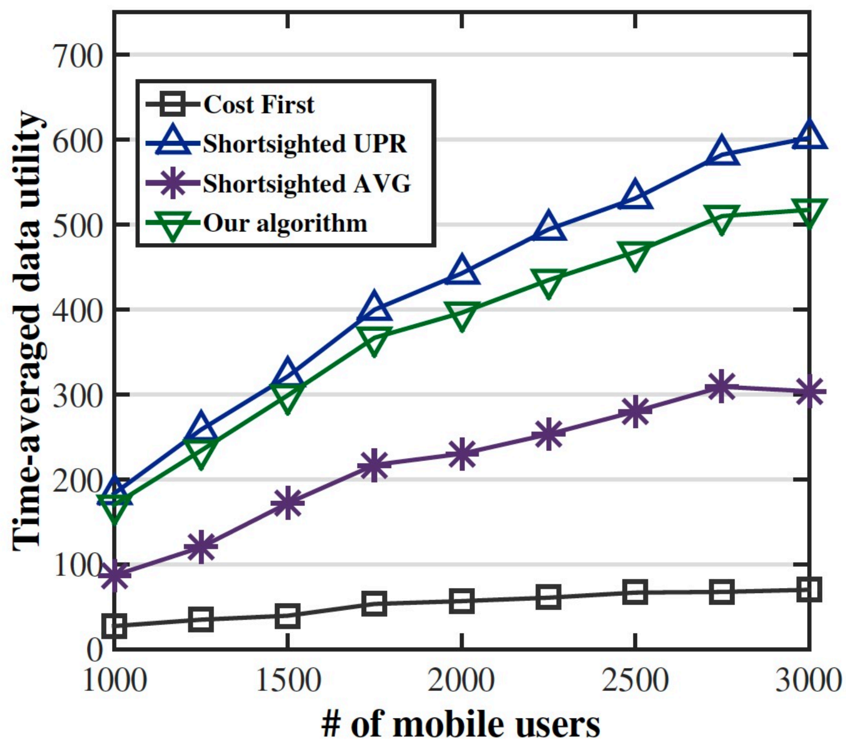


Fig. 5. Time-averaged data utility vs. number of mobile users.

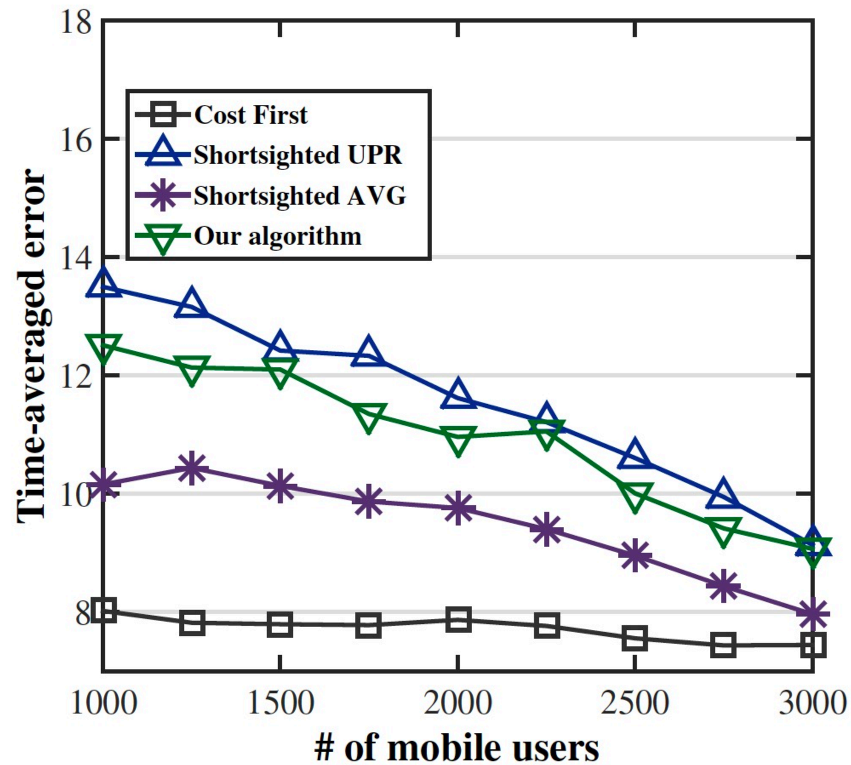


Fig. 6. Time-averaged error vs. number of mobile users.

Experiments

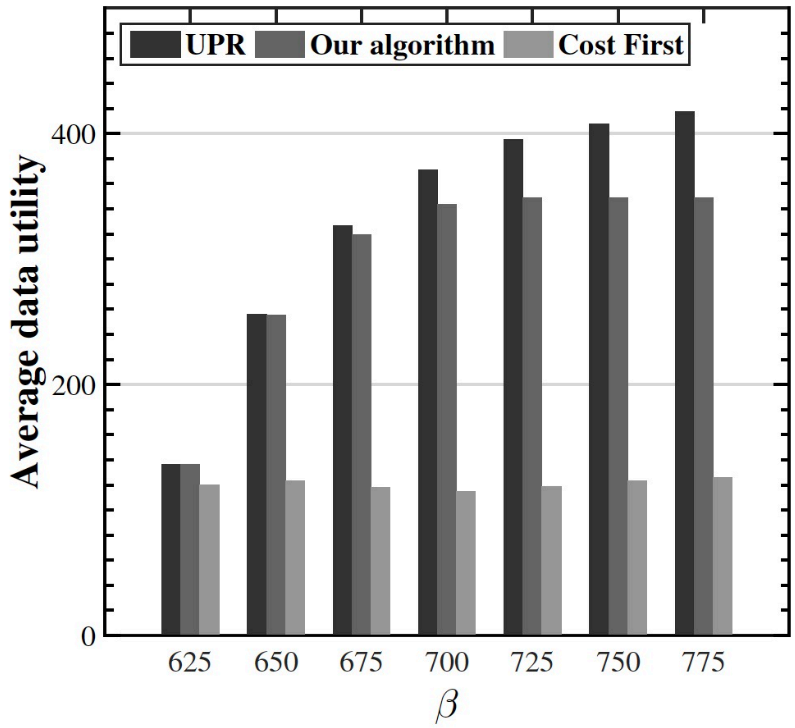


Fig. 9. Time-averaged data utility vs. upper bound β of the budget.

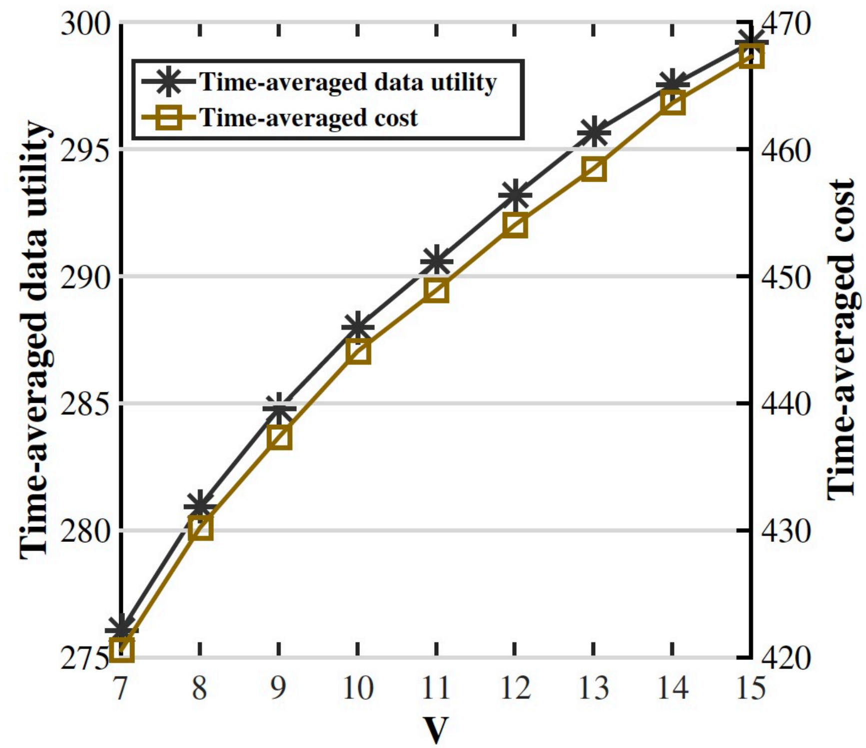


Fig. 11. Time-averaged data utility and time-averaged cost vs. V.

Road Map

- Introduction
- Model and Formulation
- Algorithm Design
- Experiments
- Conclusion

Conclusion

- In this paper, we have studied the data utility maximization problem under the average budget constraint in environmental monitoring.
- We come up with a novel data utility model to measure how good a set of data is.
- We further design an online algorithm to solve the long-term problem.