RFID Data Cleansing by Bayesian Filtering

Zhitao Gong

November 26, 2016

Outline

Introduction

System

Experiment

Discussion

Motivation

Goal – Cleanse noisy RFID readings for indoor location service.

- People spend lots of time indoor.
- GPS is unavailable.
- RFID receiver is cheaper
- RFID raw readings are noisy, e.g., false negative readings.
- Existing solutions do not fit our problem, e.g., static objects [Bab+14], grid-layout setting [Gen+14], event stream queries [Ré+08; Wel+08], etc.

Spatial Queries

Conventional definition

Range query Decide the objects that lie in a given range.

k-nearest neighbor query Decide the k nearest objects w.r.t a given query point.

Probabilistic interpretation

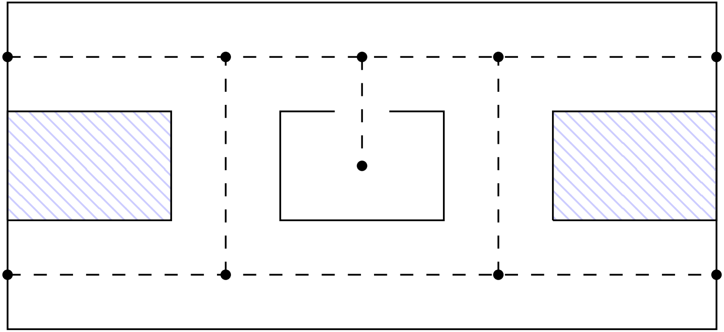
Range query Decide the objects that lie in a given range with probability larger than a threshold.

k-nearest neighbor query Decide the k' nearest objects whose probability sum is no less than k.

Simplification – Walking Graph

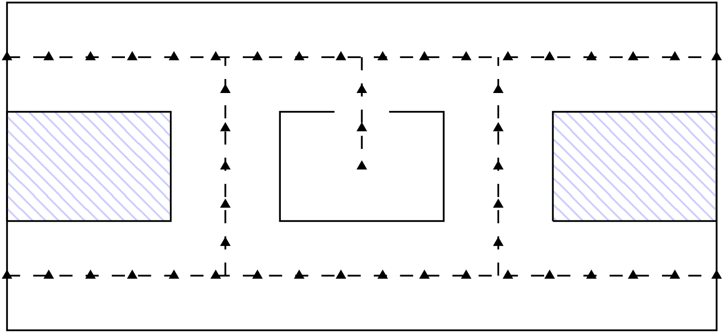
1. Hallways $\rightarrow$ graph edges.
2. Crosses $\rightarrow$ graph nodes.
3. Rooms $\rightarrow$ graph nodes.

Rooms differ from crosses: *maximum stay duration*.



Simplification – Anchor Point

- Objects are rounded to the nearest *anchor point*.
- Anchor points are indexed by R-tree.



Outline

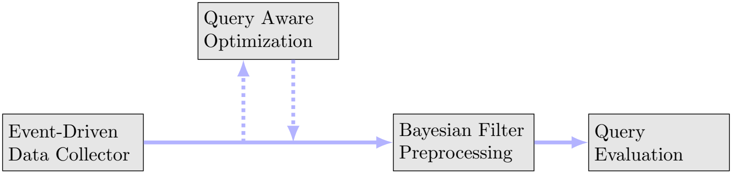
Introduction

System

Experiment

Discussion

Overview



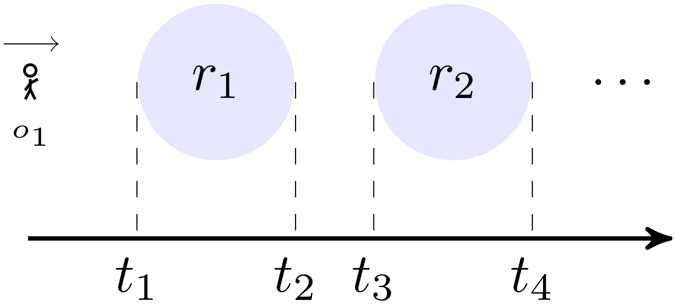
Data Collector generates ground truth data.

Query Optimization filters out *non-candidates*.

Bayesian Filters cleanse the noisy RFID data.

Query Evaluation answers spatial queries.

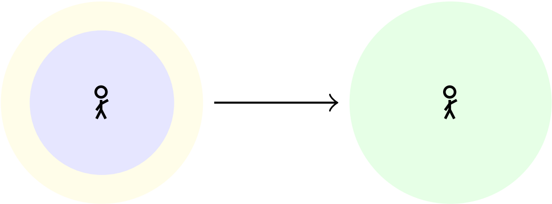
Event-Driven Data Collector



The data collected is

$$o_1 : (t_1, r_1), (t_2, \mathrm{NIL}), (t_3, r_2), (t_4, \mathrm{NIL}), \ldots$$

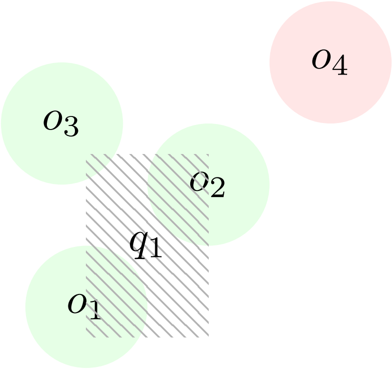
Optimization Model



We are pruning *non-candidates* by their uncertainty range.

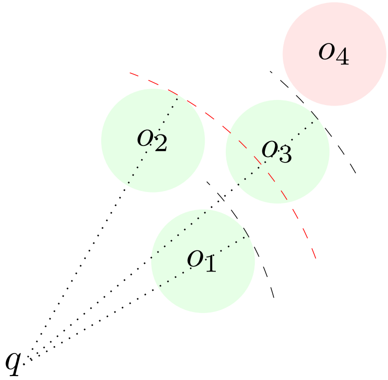
- denotes the reader range.
- denotes the uncertainty range.

Optimization Model – Range Query



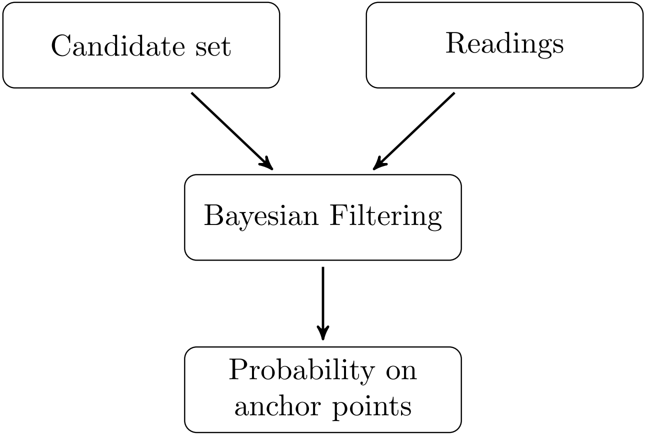
The candidate set is $\{o_1, o_2, o_3\}$

Optimization Model – *k*NN Query



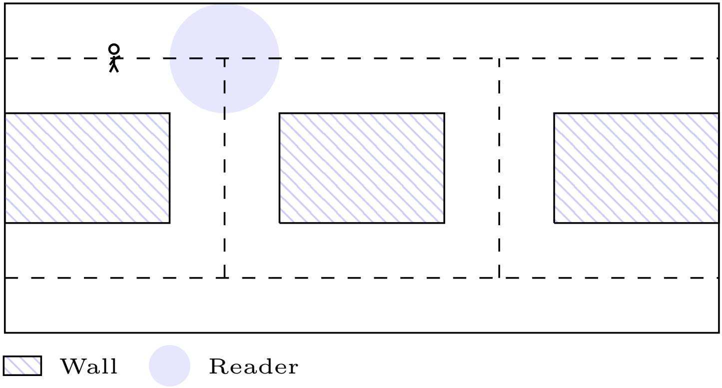
The candidate set for 2NN query $\{o_1, o_2, o_3\}$.

Bayesian Filtering



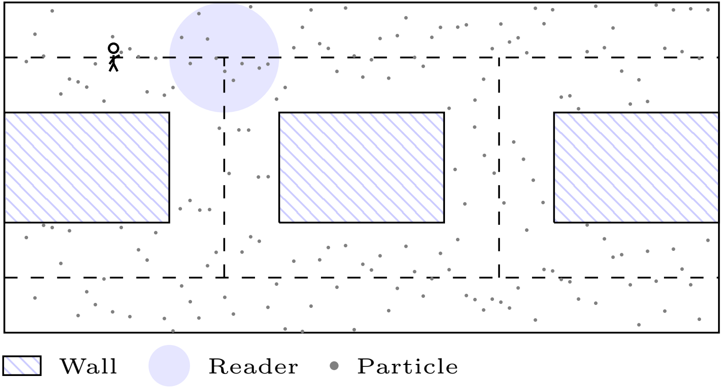
Bayesian Filtering – Particle Filter

We want to track the object.



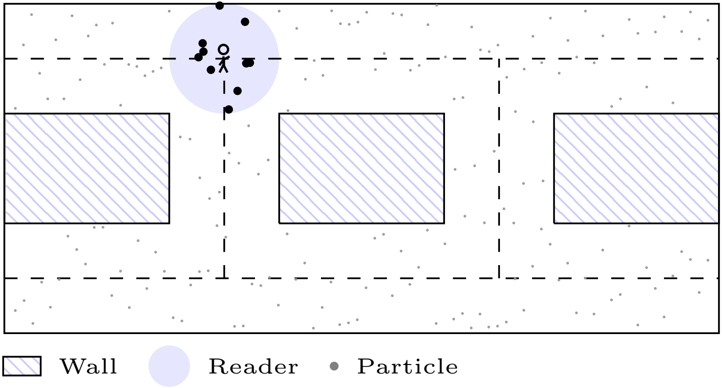
Bayesian Filtering – Particle Filter

Generate random guesses, i.e., particles.



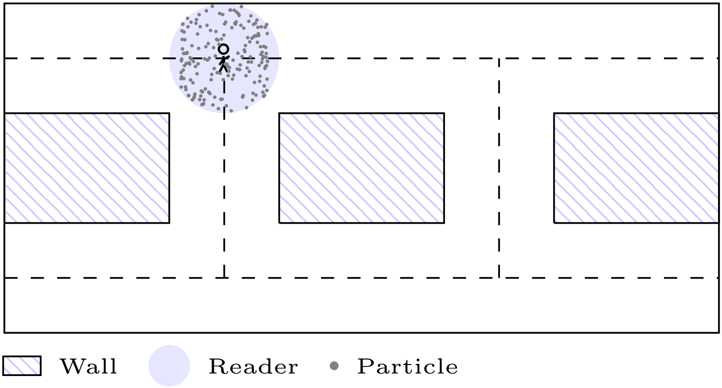
Bayesian Filtering – Particle Filter

Particles that explain the observation well get larger weights.



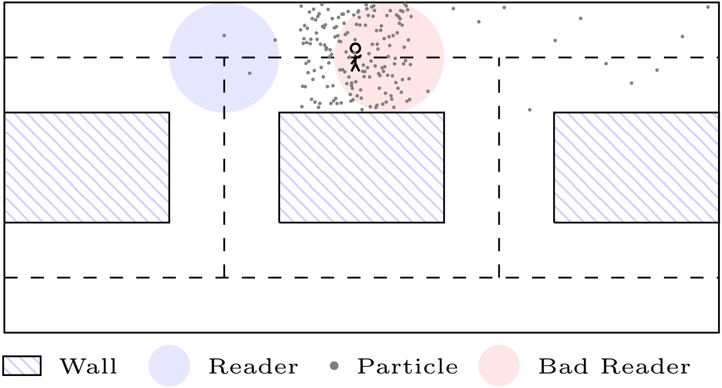
Bayesian Filtering – Particle Filter

Resample *good* particles to keep population size.



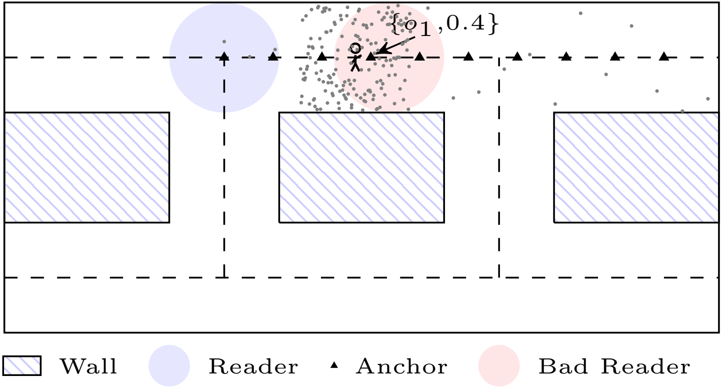
Bayesian Filtering – Particle Filter

In case of a false negative, predict with particles.



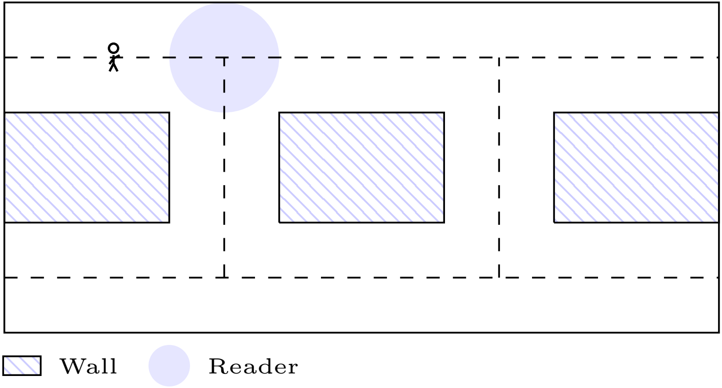
Bayesian Filtering – Particle Filter

Align particles to the nearest anchor point.



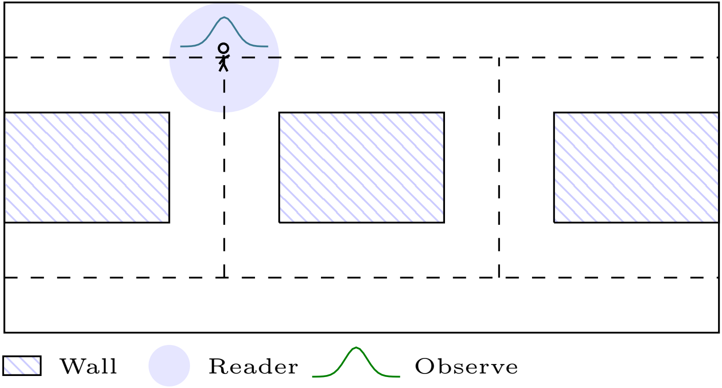
Bayesian Filtering – Kalman Filter

We want to track the object.



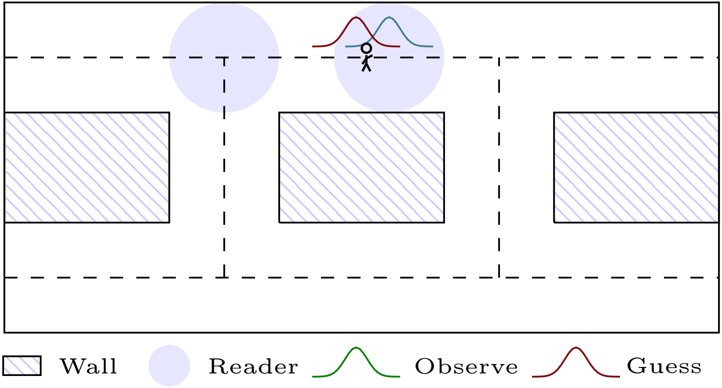
Bayesian Filtering – Kalman Filter

Each reading is an observation.



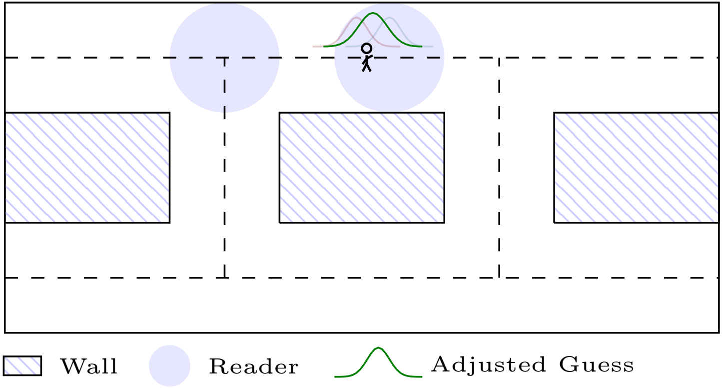
Bayesian Filtering – Kalman Filter

We have a new observation and guess from the past.



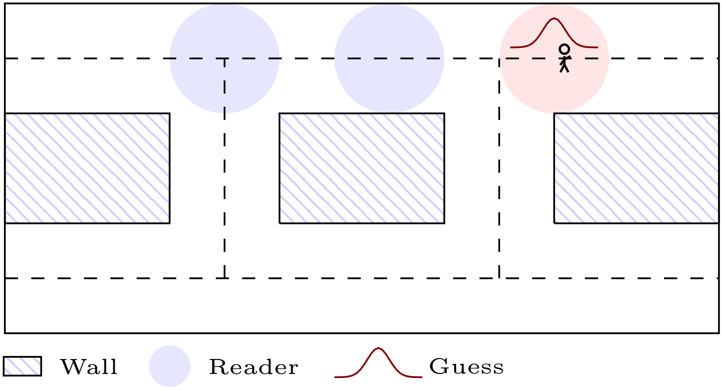
Bayesian Filtering – Kalman Filter

We combine these two to form an adjusted guess.



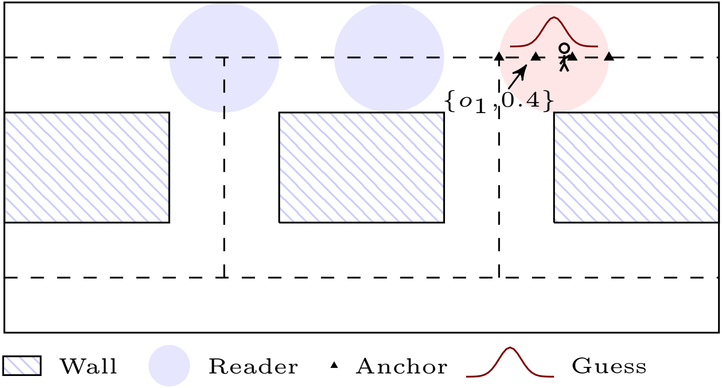
Bayesian Filtering – Kalman Filter

In case of a false negative, predict from pass guess.

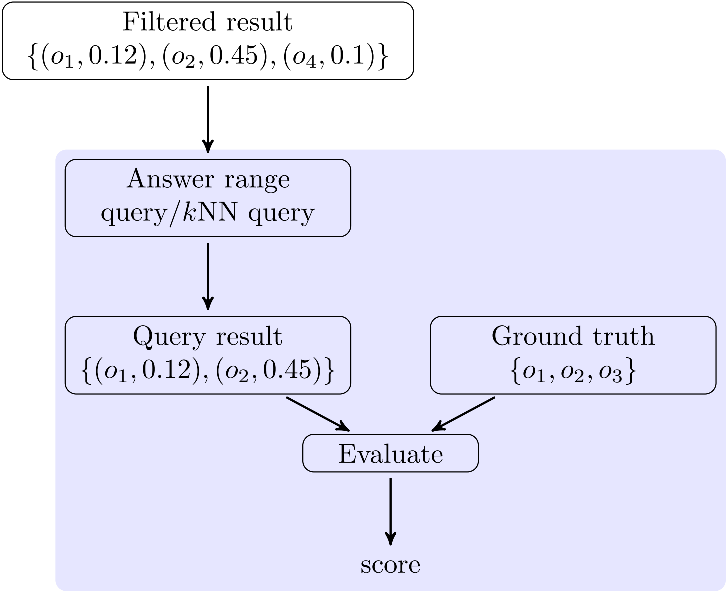


Bayesian Filtering – Kalman Filter

Assign *normalized* probability to anchors points.



Evaluation Module



Outline

Introduction

System

Experiment

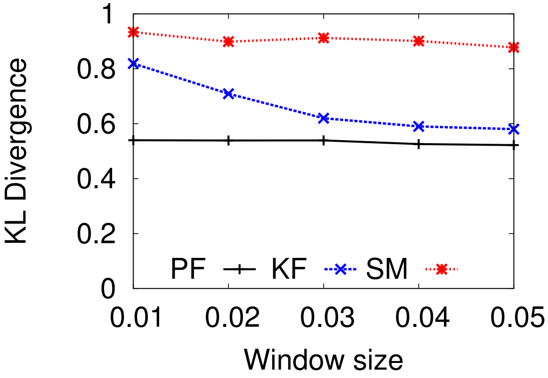
Discussion

Default Values

Parameters	Default Values
Number of particles	64
Query window size	2%
Number of moving objects	200
k	3
Activation range	2 meters

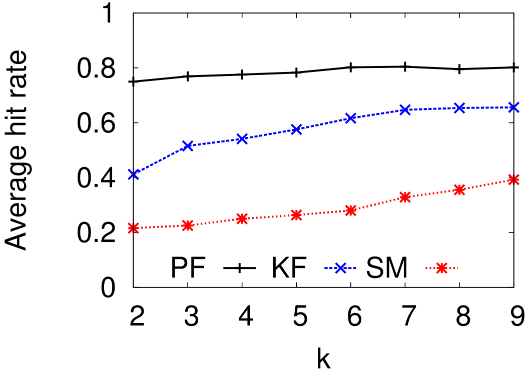
Range Query Window Size

1. KL value lower is better.
2. Query window size is the ratio of areas between query window and total environment.



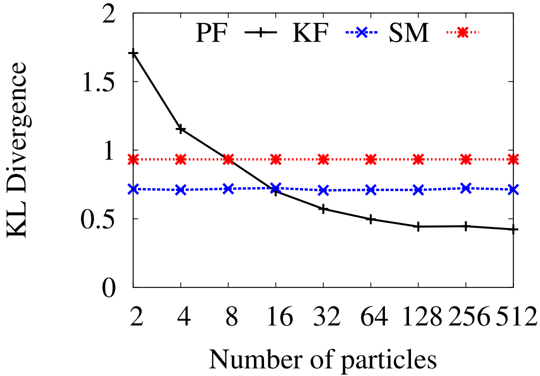
k in kNN

1. Hit rate higher is better.
2. k is the number of nearest neighbors.



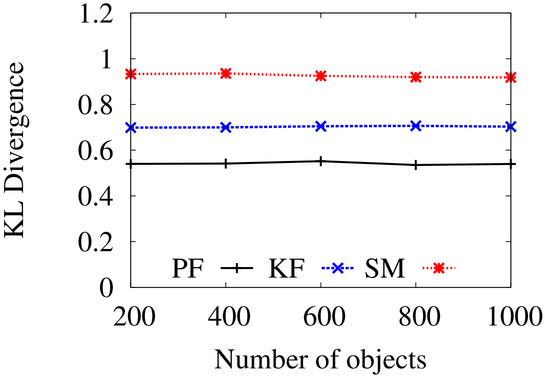
Number of Particles

1. KL value lower is better.
2. This parameter only applies to particle filter.



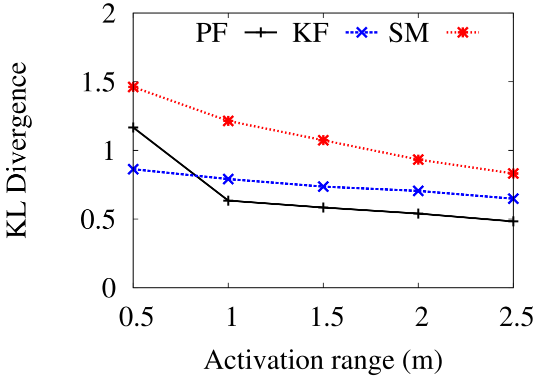
Number of Moving Objects

1. KL value lower is better.
2. Total number of objects in the environment.



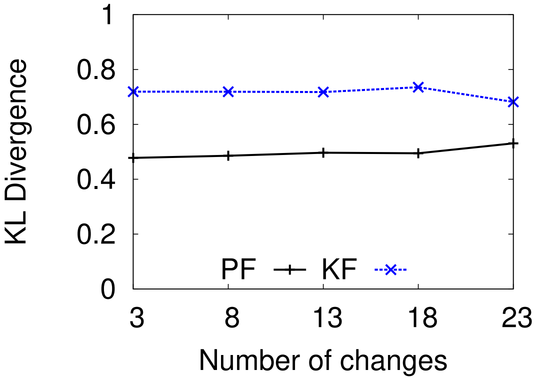
Detection Range

1. KL value lower is better.
2. The detection range of each reader.



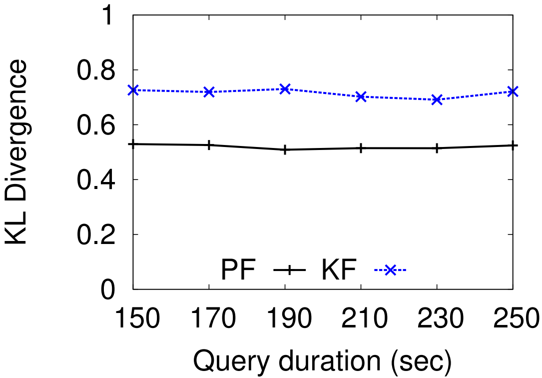
Change of Volume

Suppose t_1 the result set is $\{o_1, o_2, o_3\}$, and t_2 the result changes to $\{o_1, o_2, o_4\}$, then the change volume is 2.



Query Duration

Measure how stable our algorithm is.



Outline

Introduction

System

Experiment

Discussion

Summary

1. We designed Bayesian-filter algorithms for RFID data cleansing.
2. We propose indoor walking graph model and anchor point model to simplify the filter process.
3. We propose two metrics to evaluate continuous queries.
4. Extensive experiment demonstrates the effectiveness of our solutions compared with symbolic model.

Future Work

1. Test on real-world data.
2. Experiment with more diverse environment settings.