1 Experiments - Background

In order to test our replacement model, we conduct experiments in a controlled, simulated enviroment. The choice of our representative application is that of a mobile navigation system where a user travels on land from one location to another in his vehicle. During the trip, he asks various queries pertaining to his need and/or interests. Local cache is used to store the frequently asked queries. And replacement policy is adopted for cache management. The user juggles his time between think time and ask time. Queries are issued during the ask time only.

In order to model moving objects along with their behaviors in a 2D environment, we could use a real datasets, roads and networks (obtainable from variety of sources ... cite them). But real datasets are complex, cumbersome and difficult to manipulate. Instead we simulate the real enviroment be generating synthetic datasets to represent road networks, intersections places of interest etc. The workspace we define is a two dimensional area with area centric points defining hotels, restaurants, gas stations etc. We also need a model to emulate the mobile behavior of an object (moving vehicle). We use the existing synthetic dataset generator GSTD [PT00] for this purpose. GSTD generates sets of moving points/rectangular data that follow extended set of distributions. GSTD creates trajectories of moving objects. These trajectories could be considered as roads on which the objects move. Then, we select certain area centric points on these trajectories and overlay point data like hotels, gas stations on them. Thus we model the real life mobile navigation application in a 2D environment using synthetic datasets.

There are other techniques like

Advantages of using GSTD

Other alternatives
City Simulator simulates objects moving in a city o/p x,y coordinates, timestamp

customizable parameters people observation points probability (entry , exit in the building, drift on the road, speed)

problem the parameters shange when a certain threshhold on the number of people in the building is changed. We do not have control over the parameters after that.

2. pseudo trajectory real network - chicago

GSTD creates trajectories of moving objects. We could consider these trajectories as roads on which objects move. Then, we could select certain area centric points on these trajectories and overlay point data like hotels, gas stations on them. These could be randomly overlaid or distributed using some distribution like gaussian.

2 Experimental Setup

For our experiments we need two things:
Dataset
Query workload

Dataset:
Motion of a real world entity is guided by its goals and hence its behavior is never completely random. However, considering infinite moving objects with their varying daily goals, we assume the overall behavior of the entire system in totality to be random. Hence we argue that a synthetic data generator producing completely random trajectories for moving objects will be a reasonable model to represent the motion of the real world objects in our application domain. We have selected the GSTD [PT00] model for this purpose. GSTD generates sets of moving points/rectangular data that follow extended set of distributions (random/gaussian/skewed).

We assume that all objects move in a 2D space within the area defined by the user. Controllable parameters of the dataset are:

- Type of object: Point object.
- Cardinality of moving object: Since our main emphasis is on semantic caching, we select a single

moving object for this purpose.

- Speed of moving object: We can specify a range for minimum and maximum speed.
- Direction of motion: We can chose the direction of motion (north/south/east/west/random).
- Initial position of the object.
- Number of snapshots: Number of observations that can be made during the run of the experiment.
- Frequency of updates: The frequency at which the observation regarding the moving object (current-location,current-time) is recorded.

Each value is calculated using a random generator.

Just for the sake of simplicity, we consider a single relation with atleast the following attributes defined: (Location-x, Location-y, Hotel-Name/School-Name, Hotel-ID/School-ID)
Query workload selection parameters are:

- Type of queries: Rectangular queries (Select and Project operations only)
- Location Dependence: Query predicates involving the current location of the moving object.
- Temporal/Prediction queries (Optional): Queries involving time as one of the predicates
- Selectivity of results: Number of tuples present in the result. We can place a max cap on the number of tuples fetched, depending on the size of the cache.
- Overlap rate: If there is no overlap between the queries (and hence their corresponding results), we would not derive any benefits from caching. We define the overlap rate to be the rate at which the queries overlap each other. We could set the value, for example 25%, meaning, on an average every fourth query overlaps with one or more previous queries.
- Think time: The time between two consecutive queries issued by the user. We set this value in a range from a minute to five minutes.

References

[PT00] D. Pfoser and Y. Theodoridis. Generating semantics-based trajectories of moving objects, 2000.