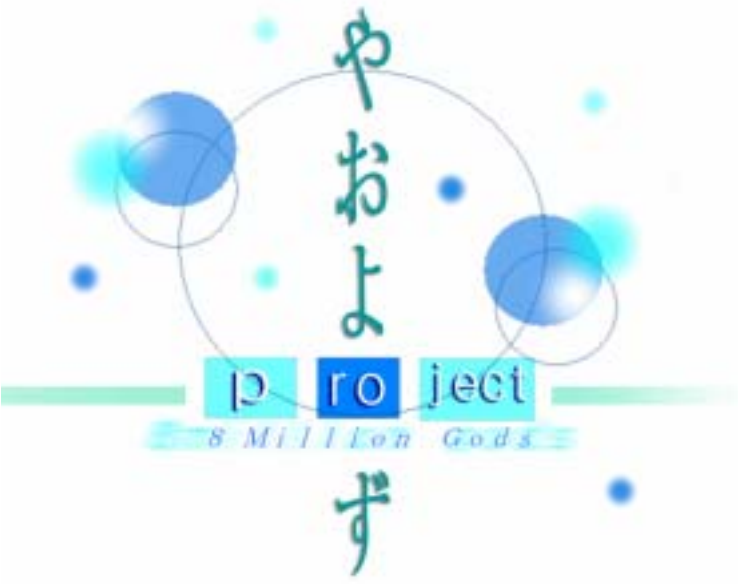


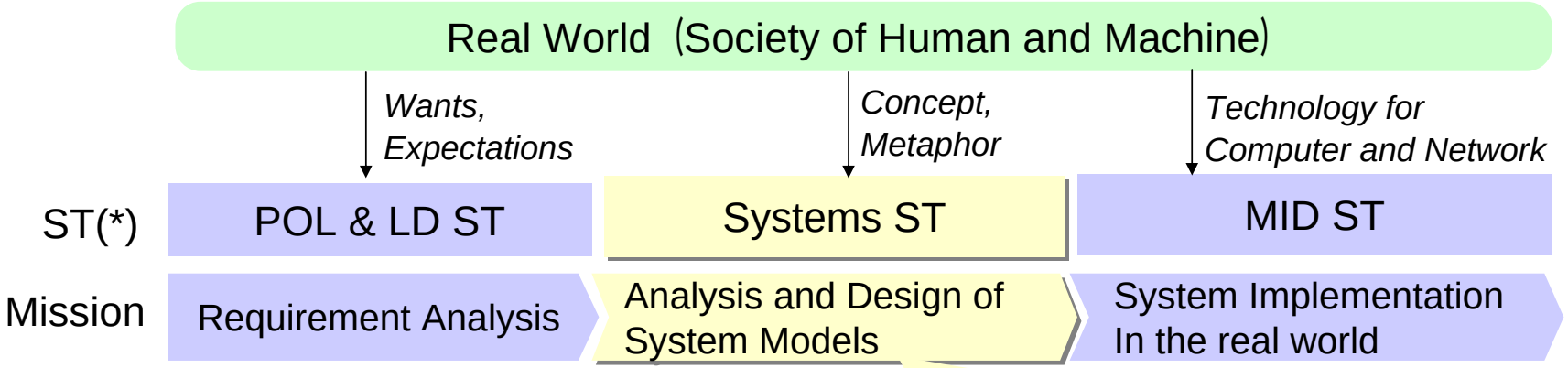
Information Filtering System based on Spatial Patterns for GPS-Enabled Mobile Terminals

Members of Systems Sub-Theme in the Yaoyorozu Project

○ Yoshinori Sato, Takashi Fukumoto, Hiromitsu Kato, Seiichi Shin

Contents

1.1 Missions of Systems Sub-Theme in the Project



- Systems concept of the Yaoyorozu project
- Basic information sharing technologies for
  - safety (anonymous trust propagation model)
  - scalability (P2P network using optimal paths design)
  - user adaptation (information filtering system based on spatial patterns)
- System evaluation framework (multi-agent simulation)

(*)ST: Sub-Theme

1.2 Background

- A large amount of information is available on the mobile terminals (cell phones, PDAs) in daily life.
- Mobile terminals may be still popular in the near future.
- Limitations of user interface on mobile terminals.
  - It should be simple because of poor input device.
  - Only small display can be used to show information.
- Needs for more intelligence to select appropriate information for mobile terminals.
  - How to estimate user's situation in daily life.
  - How to decide priority of retrieved information for a user.

1.3 Early Works on Context-awareness for Information Filtering

Lots of works especially for location-awareness....

- Simple model
  - Some cell phone services based on GPS or base stations.
  - The Lancaster Guide project [Davies, 2001]
- Advanced model
  - Rule driven model: Ubiquitous Personalize Agent [Hattori, 2003]
  - State-sequence model: User activity assistance system [Isoda, 2003]

Problems

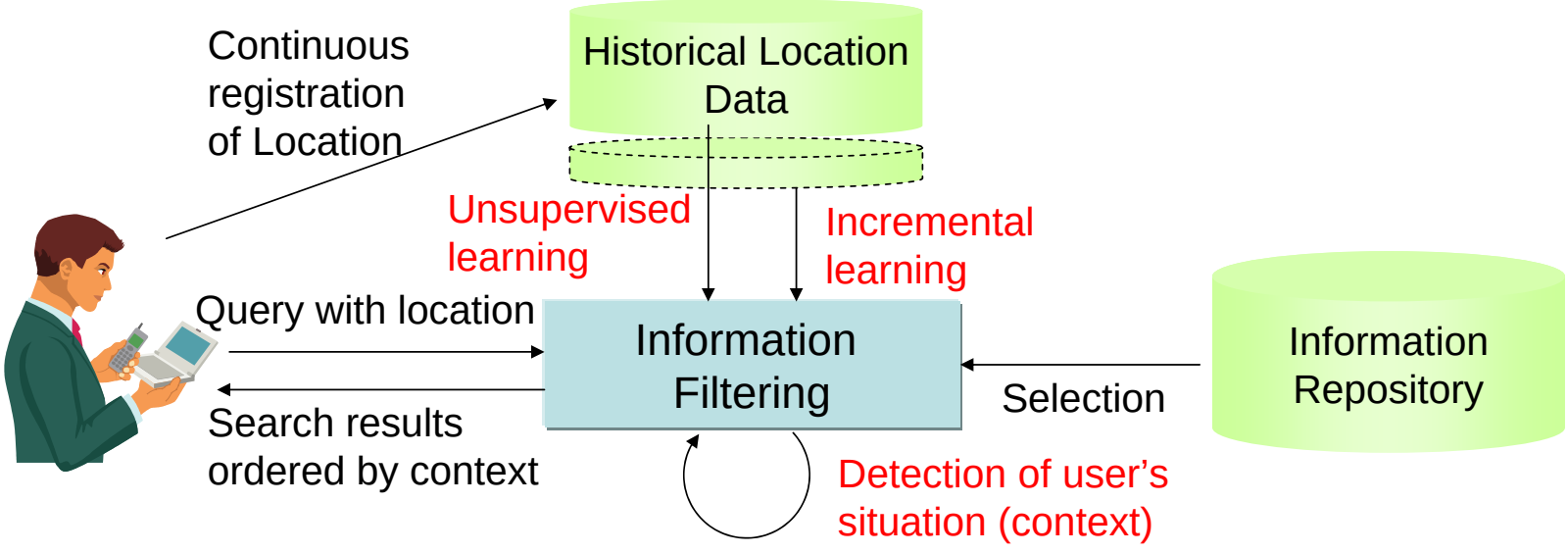
- A user may require different information at the same location. It may depend on his current task.
- Initial cost for building personal model.
- Maintenance cost for updating personal model.

GPS: Global Positioning System

2.1 Objectives

Improving context-awareness feature for Information filtering

- Detection of user's current situation that appears in spatial behavior. (long term-preferences and short-time interests)
- Automatic model construction without any feed back from a user. (unsupervised learning)
- Automatic model updating. (incremental learning)



2.2 Examples of Applications

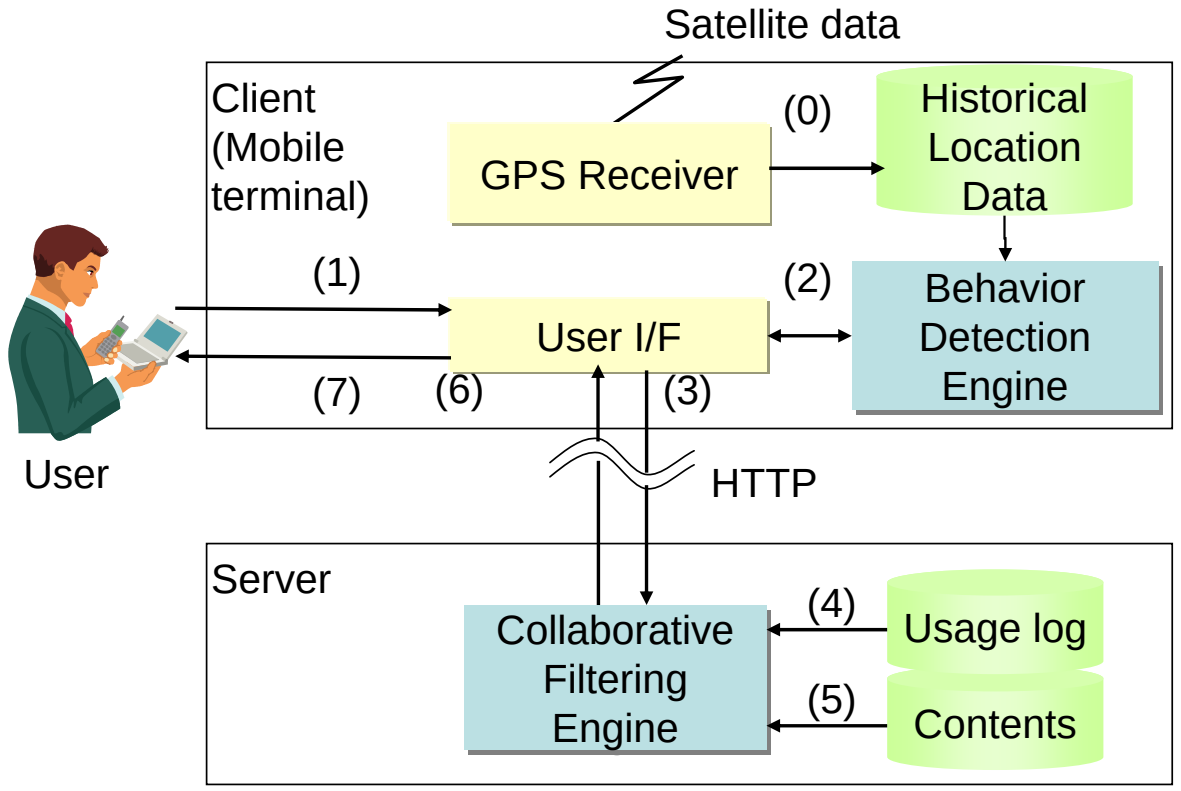
- Portal personalization
  - List contents in order of personalization for a user if he behaves normally
  - Suppress personalization if a user behaves unusually
- Banner ad targeting
  - Take account of access log of a user if he behaves normally
  - Take account of sight character that indicates banners if a user behaves unusually
- Search assist
  - Indicate candidate keywords from personal history that relate to current query
  - Indicate candidate keywords from ranking in all of users that relate to current query

2.3 Hypotheses about User's Situation in the Real World

- Users do not have a short-time interest when they behave normally, following the same pattern they usually do.
- Users have a shot-time interest when they behave spontaneously, and do something unusual.
- It is possible to observe such behavioral differences as spatial behavior pattern differences in the real world.

2.4 Information Filtering System based on Spatial Patterns

- System switches output that relates to user's short-term interests and long-term preference using historical location data.
- Behavior detection engine on the client side.
- Collaborative filtering engine on the server side.



2.5 Roles of 2 main components

- Behavior detection engine
  - Understand user's short term interests and long-term preferences.
  - Check on spatial behavior pattern of a user in the real world.
  - Typically implemented in a mobile terminal in order to be careful of privacy issues.
- Collaborative filtering engine
  - Retrieve appropriate information for a user.
  - Check correlation among users in the virtual world.
  - Typically implemented in a server in order to use evaluation of each information that voted by all users.

3.1 Approaches for Detecting User Behavior

Requirement: learning model in unsupervised and incremental manner to distinguish between normal and unusual behavior.

- Classification approach (ex. ID3, C4.5, etc.)
  - × Learn a model in supervised manner from training set.
  - × It is difficult to update a model incrementally.
- Clustering approach (ex. hierarchical methods, k-means, etc)
  - ○ Learn a model in unsupervised manner from training set.
  - × It is difficult to determine which subset data relates to normal behavior or unusual behavior.
  - × It is difficult to update a model incrementally.
- Model-less approach (ex. Memory-based Reasoning)
  - ○ Pre-learning is unnecessary (It directly compares recent data with past ones).
  - ○ It is easy to update a model incrementally.
  - × Time complexity of process on searching stage is not good.

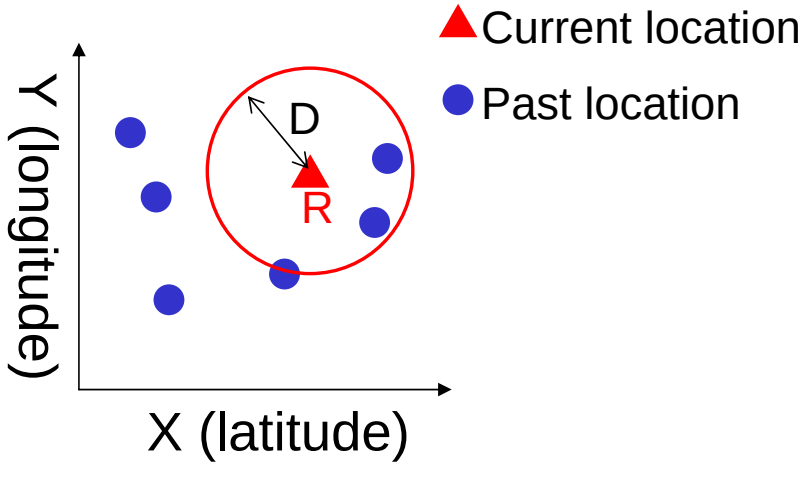
- Region based detection (RBD)

  If a user has visited location *R* more than *F* times, his behavior is normal; otherwise, it is unusual.

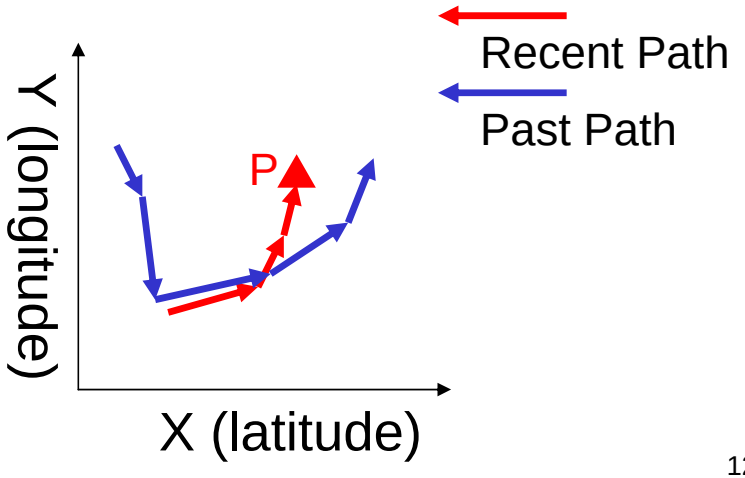
- Pass based detection (PBD)

  If a user has traced recent path *P* more than *F* times, his behavior is normal; otherwise it is unusual.

RBD Algorithm

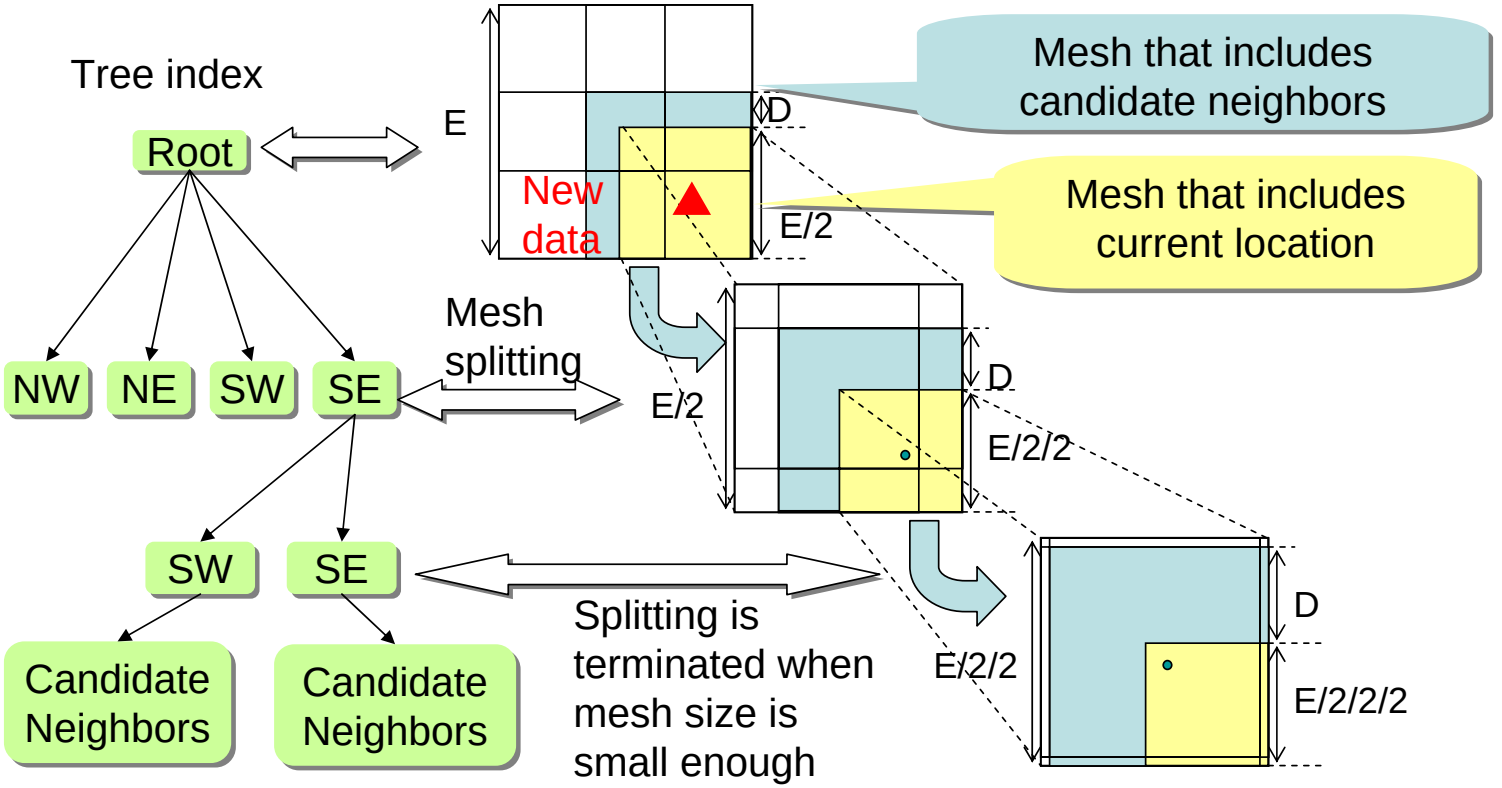


PBD Algorithm



3.3 Acceleration for PBD & RBD Algorithm

- Bottleneck: Searching neighbors of current location
- Approach: Spatial Indexing (Overlapped Quadtree)
  - In particular for geographic data (2 dimensional data)
  - Time consumption of searching is stable.



4.1 Collaborative Filtering

- A popular technique for information retrieval introduced by Rensick, P (1994).
- It locates users who are similar to the target user.
- Major application :recommenders of shopping site on the Web.

Target User A

I have voted contents α and β.

Collaborative Filtering

Users who voted α and β, also voted γ

Calculate similarity among users.

Confidence Matrix (User vs. Contents)

	Contents α	β	γ	...
User A	1	1	0	...
User B	0	1	0	...
User C	1	1	1	...
....	....	....	....	...

0: Not voted
1: Voted

4.2 Coupling Behavior Detection with Collaborative Filtering

- Behavior detection engine switches source data about the target user that compiled to confidence matrix.
- Compiles long-range data if user behaves normally.
- Compiles only recent data if user behave unusually.

Source data about the target user A

Date	Time	Contents
Apr. 1, 2005	10:00	α
Apr. 3, 2005	10:02	α
....	...	...
May 1, 2005	12:00	β
May 1, 2005	12:15	β

Time

Switch compiled data

Confidence matrix

	Contents α	β	γ	...
User A				
User B	0	1	0	...
User C	1	1	1	...
....	....	....	....	...

5.1 Experimental Conditions

- Benchmark target : Time consumption of RBD algorithm
  - Naive method vs. Indexed method (Overlapped quadtree)
  - Find the nearest neighbors within 1 km from the current location
- Platform (PDA)
  - CPU: 400 MHz
  - Memory: 128 MB
  - OS: Windows Mobile 2003
- Test data
  - Past location : randomly generated
  - Current location :3 points that randomly selected for each test.

5.2 Experimental Results

- Indexed method finished detection within 1 second in cases where the amount of data was 10^7
- Notice that the amount of data in location database will be nearly 1.5×10^7, if the system measures the user location each 1 second during tracking.

Average Time Consumption

	Amount stored in Location Database				
	10^4	10^5	10^6	10^7	10^8
Naive method	41	421	4,196	41,669	415,850
Indexed method	1	6	62	630	6,252

Units: milliseconds

6. Conclusions and Remarks

- A Framework of information filtering that appropriate the ubiquitous information environment.
  - Understands normal and unusual behavior of a user
  - Switches focusing to short-time interests and long-term preference.
- Algorithms for behavior detection from GPS data.
  - PBD & RBD algorithm in model-less approach
  - Overlapped Quadtree as spatial indexing for geographic
- Preliminary experimental results indicate potential performance for the real time processing to handle with a large amount of location data.
- Future issues
  - Evaluation of accuracy of behavior detection.
  - Feasible study in realistic application.

Thank you for your attention.

This research is part of the “Promotion of Leading Research” program and was supported in part by a grant from the Special Coordination Funds for Promoting Science and Technology, provided by the Ministry of Education, Culture, Sports, Science and Technology.