Simulating large-scale participatory sensing data from representative set and its use in rapid transit system

Rachit Agarwal
Project-team: MiMove
Associate team: Sarathi
Inria Paris

June 9, 2016

Big picture

- Participatory sensing data
  - Time consuming to gather Large scale data
- Recommendation Systems face cold start problem
- How could we use limited data gathered to provide recommendations
- Use Case: Rapid Transit System

© Consuelo Barreto,123RF

Objective

- How to simulate large-scale participatory sensing data from representative set (already existing limited size data) and use it to provide personalized recommendations towards Rapid Transit System

Participatory sensing and its Challenges

- Participatory Sensing[1] "is the concept of communities (or other groups of people) contributing sensory information to form a body of knowledge" using mobile devices.

- Challenges:
  - Time: May take upto ~~~~ weeks ~~~ de ~~~~ of data
  - Lack of user participa~~~~ ~~~tivatin~~ ~~~peop~~~ de data
  - Correctness of data, quality of data, Privacy of data, etc.

- Question: How to motivate users to submit data that can be used towards achieving the desired goals?

[1] "Participatory sensing", https://en.wikipedia.org/wiki/Participatory_sensing

Motivating Users

- Provide **incentives** which are :

    -

    - may not always motivate user participation

    - Altruistic: e.g., personal satisfaction

    - Democratic: e.g., helping the community

[2] Yang et al., "Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing", MobiCom 2012
[3] Lee et al., "Sell your experiences: A market mechanism based incentive for participatory sensing", PerCom 2010
[4] Zhang et al., "Reputation-based Incentive Protocols in Crowdsourcing Applications", InfoCom 2012
[5] Rashid et al., "Motivating Participation by Displaying the Value of Contribution", CHI 2006
[6] Kawasaki et al., "Top of Worlds: Method for Improving Motivation to Participate in Sensing Services", UbiComp 2012
[7] Lopez et al., "Towards Adaptive Recruitment and Engagement Mechanisms in Social Systems", UMAP 2011

A Solution

- Rely on the available collected dataset
    - Assume: it is representative
- Identify properties and patterns of within the collected dataset
- Simulate data using identified properties and parameters to create larger dataset and negate cold-start problem in training recommendation system
- Use the new simulated dataset to recommend
- Evolve the recommendation when actual data is available

UseCase: Rapid Transit System

- Personalized recommendations towards convenient metro routes to commuters

- Knowledge gathered:

n legs

$$\{submissionTime, source, destination, userID,$$
$$line_1, comfort_1, delay_1, seat_1,\ldots,$$
$$line_n, comfort_n, delay_n, seat_n,$$
$$ratingOverall\}$$

- Please install the app and help us "Democratically"

Available for download at:
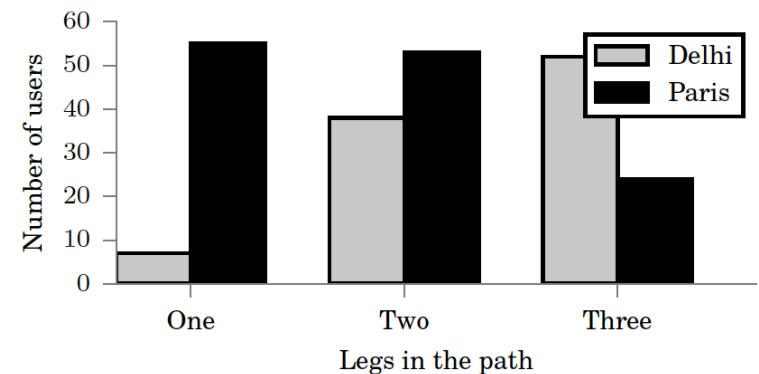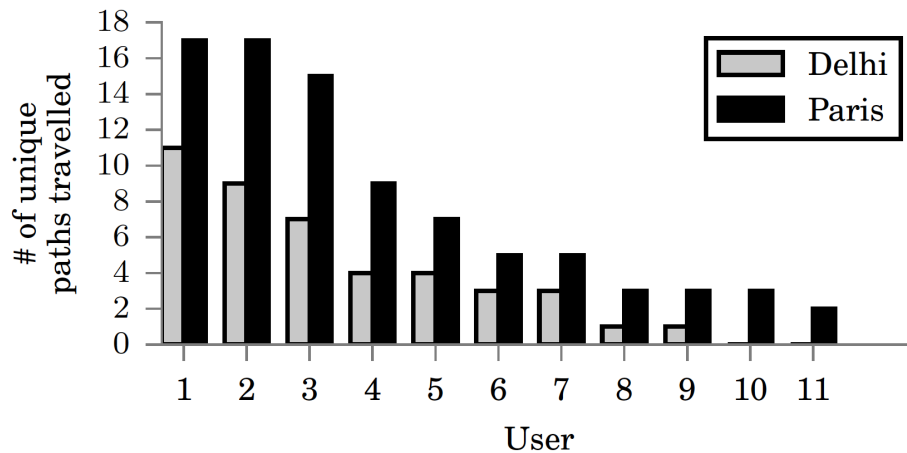https://play.google.com/apps/testing/edu.sarathi.metroCognition

Gathered Dataset: Properties and Patterns

{submissionTime, source, destination, userID, $line_1$, $comfort_1$, $delay_1$, $seat_1$,…, $line_n$, $comfort_n$, $delay_n$, $seat_n$, ratingOverall}

- Small set of available users provided ratings

Property	Delhi	Paris
Number of users	9	11
Number of unique paths rated	37 (total 98)	77 (total 77)
Number of paths	10^{19}	10^{33}
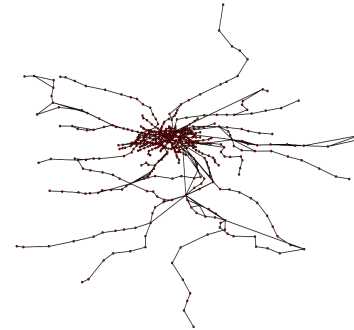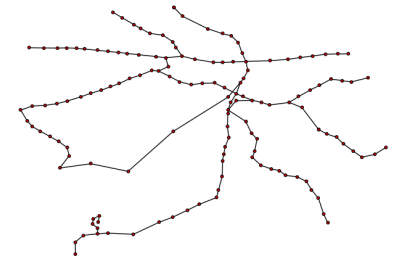
Gathered Dataset: Properties and Patterns

{submissionTime, source, destination, userID, $line_1$, $comfort_1$, $delay_1$, $seat_1$,…, $line_n$, $comfort_n$, $delay_n$, $seat_n$, ratingOverall}
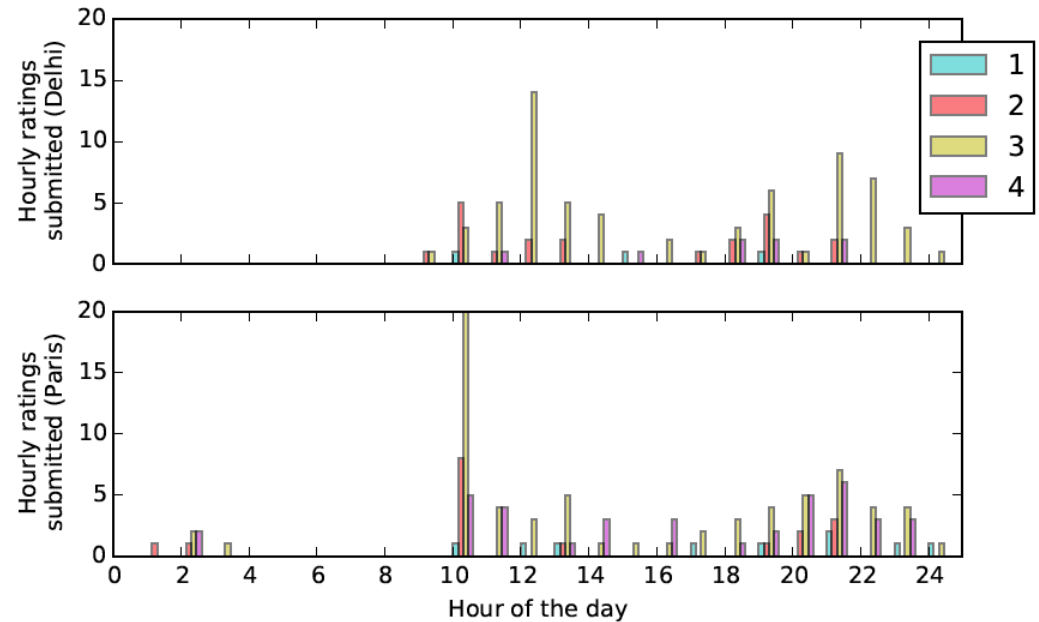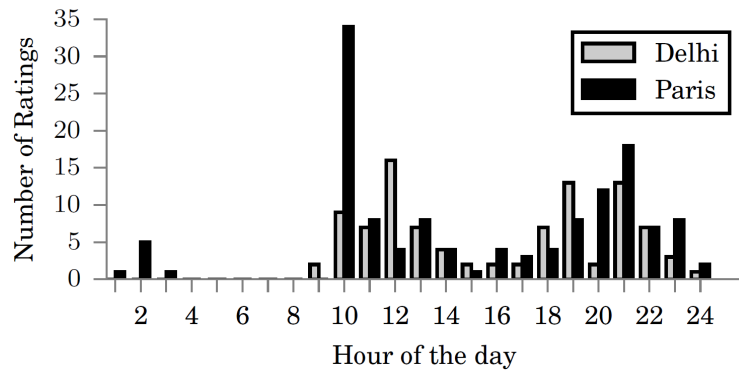
- Properties and patterns identified for the data gathered:
 - Mobility

Gathered Dataset: Properties and Patterns

{submissionTime, source, destination, userID, $line_1$, $comfort_1$, $delay_1$, $seat_1$,…, $line_n$, $comfort_n$, $delay_n$, $seat_n$, ratingOverall}

- Properties and patterns identified for the data gathered:

    - Rating Patterns

Data Simulation: one approach

- 1000 users, 1000 paths per user (unique paths form subset)
- Randomly select 'home' and 'office' locations
  - Assumption: human mobility is predictible and follows patterns
    40% paths – home to office (morning hours)
    40% paths – office to home (evening hours)
    20% paths – randomly selected
- Ratings assigned based on the probabilities of rating in an hour
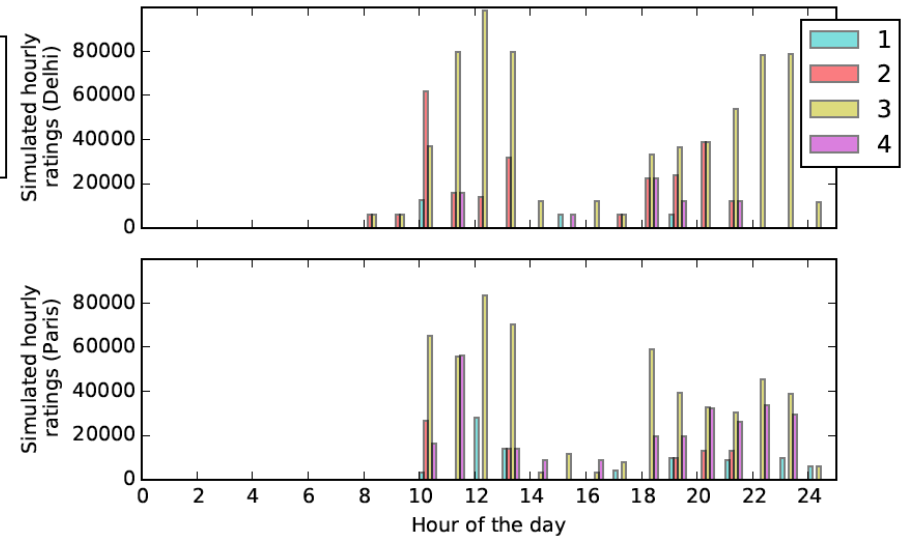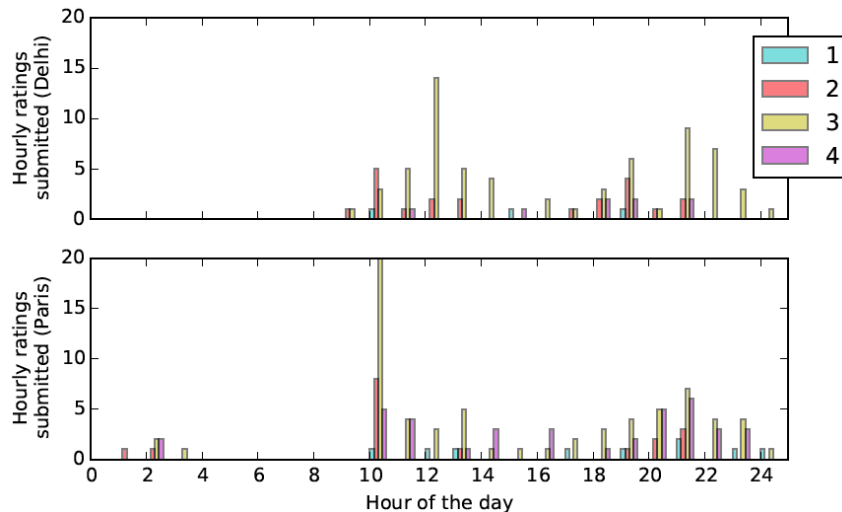
Studied Recommendation System

- User-based collaborative filtering using k-NN
- Probabilistic Matrix Factorization (PMF)
- Bayesian Probabilistic Tensor Factorization (BPTF)

	User-based		PMF	BPTF
	Pearson (k=30)	Euclidean (k=100)	(D=10)	(D=10)
RMSE(Paris)	0.918	0.856	0.951	0.752

	User-based		PMF	BPTF
	Pearson (k=30)	Euclidean (k=100)	(D=30)	(D=30)
RMSE(Delhi)	0.7566	0.6412	0.6182	0.5430

Conclusion

- Preliminary work
- Representative data can be used to provide personalized recommendations

Future Work

- Identify more properties
- Validatation
- Gereralization of the approach
  - To dataset (Modality, etc..) wherever possible
- Quantify representative set
- More fine grain study

Thank You

Email:
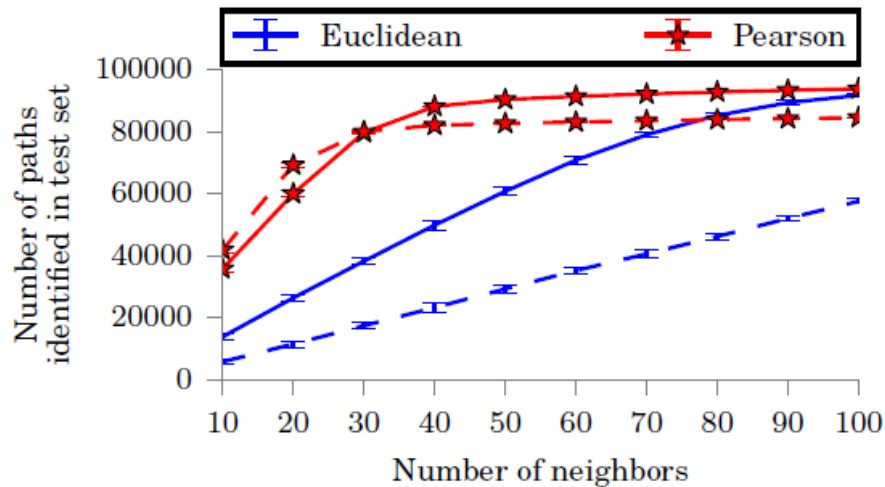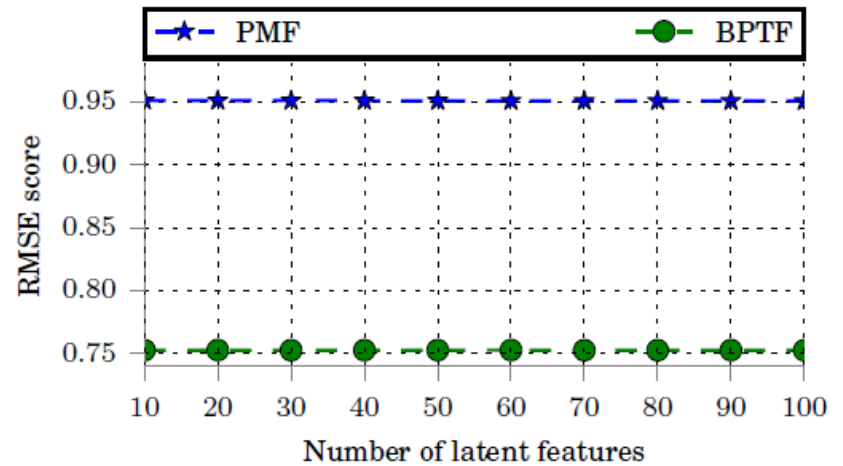
rachit.agarwal@inria.fr

https://sarathi.gitlab.io/web/

User-based collaborative filtering

- Number of neighbours : 10 to 100
- 15 iterations

PMF and BPTF

Who Am I