Movement Prediction

Amirabbas Jalali, Mina Mousavifar

Motivation

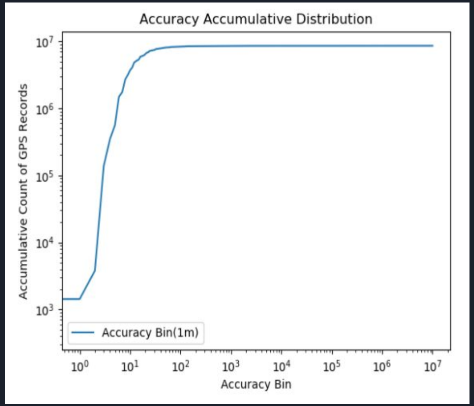
- Investigating the possibility of predicting user locations based on his/her location history
- Fill the users’ data gaps to preserve more users in the dataset for analyzing.
- Applying imputation methods to obtain location prediction speedily and specifically

Data

- Saskatchewan Human Ethology Datasets (SHED10)
- 108 participants
- February 6th to March 7th, 2017 (28 days)
- 5-minute duty cycle
- 8,592,409 GPS records

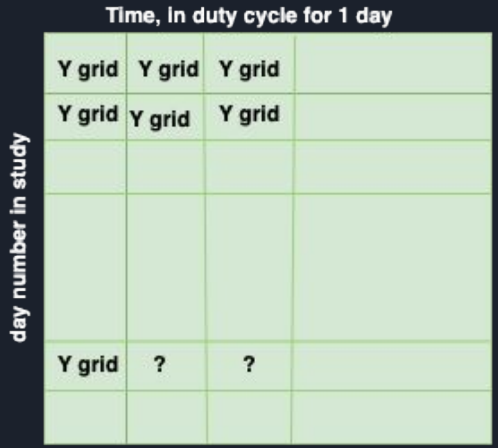
Filtering, Aggregation, Stratification

- **Filtering**
  - Filter users with more than 50% battery records - 42 users
  - Filter records with less than 100m accuracy - 97.46% GPS records
  - Limit latitude and longitude to Saskatoon
- **Stratify and Aggregate**
  - Extract 5 minutes duty cycle for records
  - Aggeragate latitude and longitude over duty cycles
  - Convert coordinates to Universal Transverse Mercator (UTM)
  - Use 100m grid size

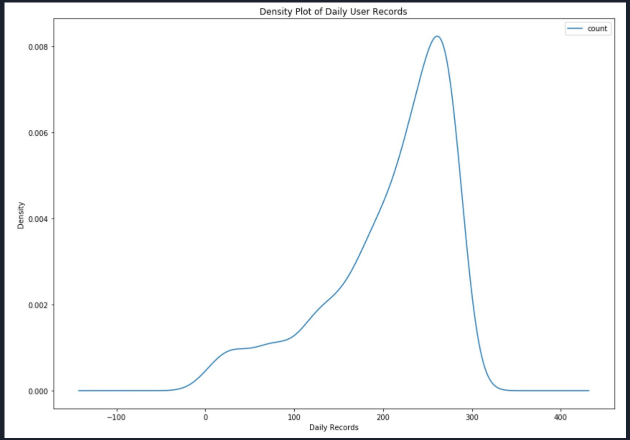


User Placement Matrix

- Users’ placement matrix expresses the user location (grid cell) every 5 minutes
- Each row represents a day in the experiment
- Each column represents five minutes time spans in a day
- Create interaction matrix for each user and X and Y coordinate separately

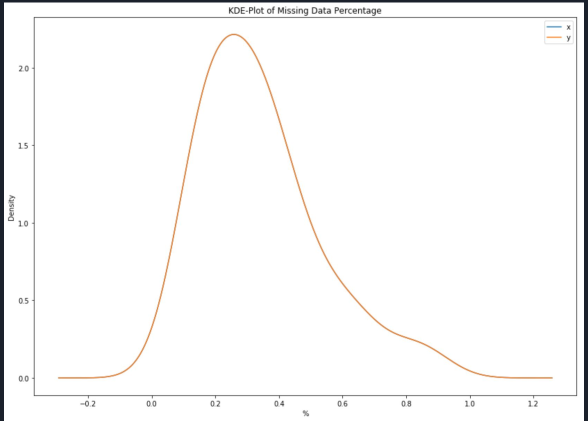


Statistics: Daily records



Summary of Grid

	Min	Max	Mean	STD
Horizontal Grid	0	165	90.7	23.9
Vertical Grid	0	173	75.5	14.3



- Randomly mask 20% of cells for evaluation

Models: Simple Fill

Imputation is the process of replacing missing data with an estimated value based on other available information

- Replacing each missing cell with the mean of each column based on the existing data
- Baseline

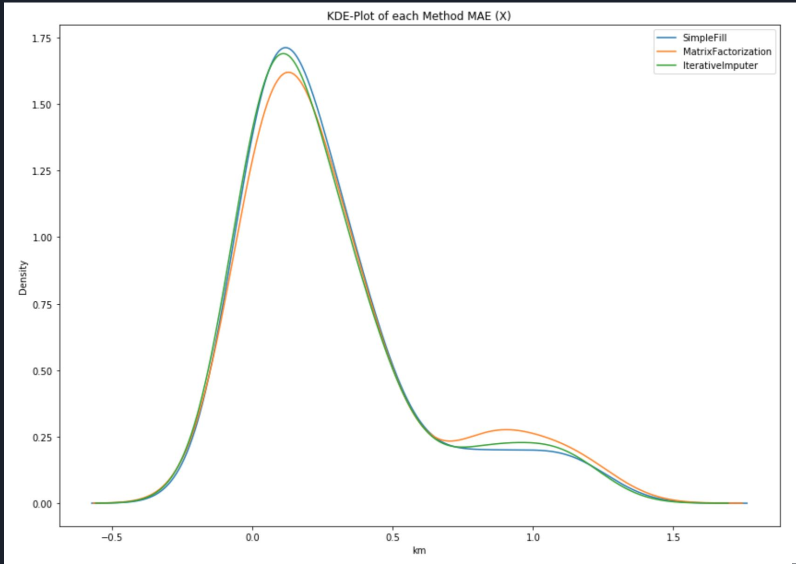
Models: Matrix Factorization

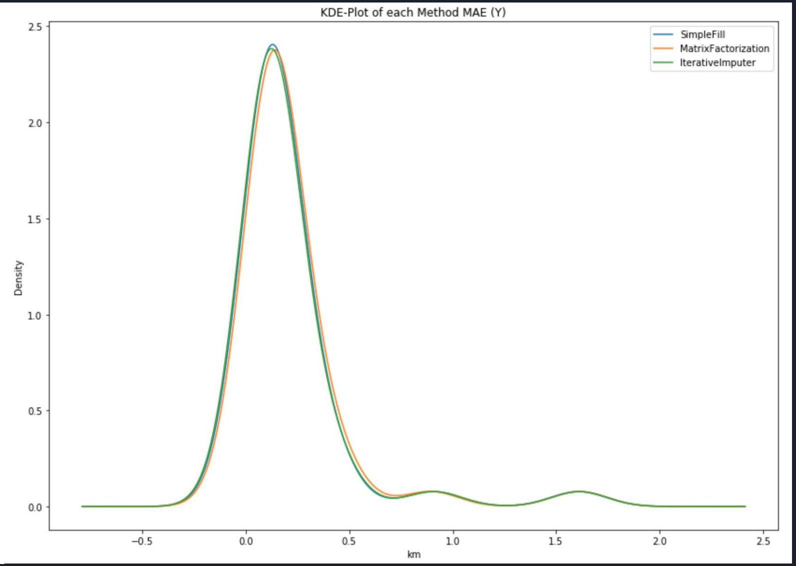
- Projects days and duty cycles into a shared latent space, using a vector of latent features to represent a day or a 5 minute span
- User placement in a day on a specific time span is modelled as the inner product of their latent vectors.
- Factorize user placement matrix using gradient descent

Models: Iterative Imputer

- Imputing missing values by modelling each feature with missing values as a function of other features in a round-robin fashion
- At each step, a feature column is designated as output y and the other feature columns are treated as inputs X
- A regressor is fit on (X, y) for known y
- Then, the regressor is used to predict the missing values of y
- This is done for each feature in an iterative fashion

Results
Masking data for one day





Horizontal Axis

Vertical Axis

Results
Masking randomly vs Masking one day

- Masking cells randomly

Model	Horizontal MAE	Vertical MAE
Simple Fill	306.2235 m	49.2045 m
Matrix Factorization	220.7322 m	41.1324 m
Iterative Imputer	147.20205 m	31.0712 m

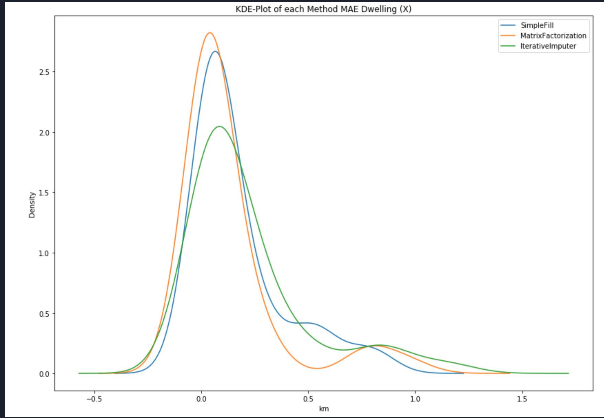
- Masking for one day

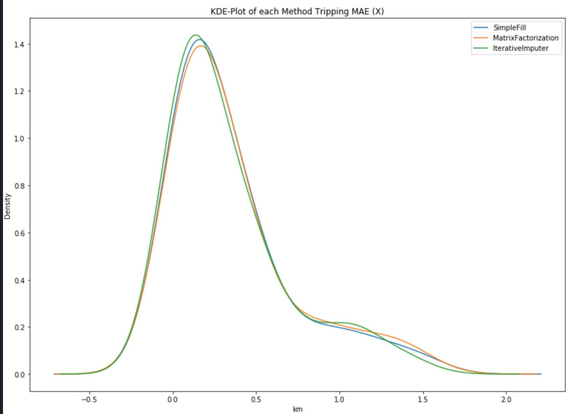
Simple Fill	271.5531 m	208.6263 m
Matrix Factorization	296.2389 m	220.7996 m
Iterative Imputer	270.1739 m	205.3427 m

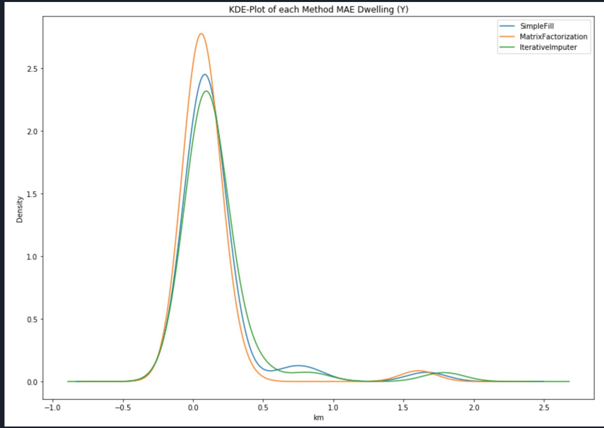
Results: Dwells vs Trips

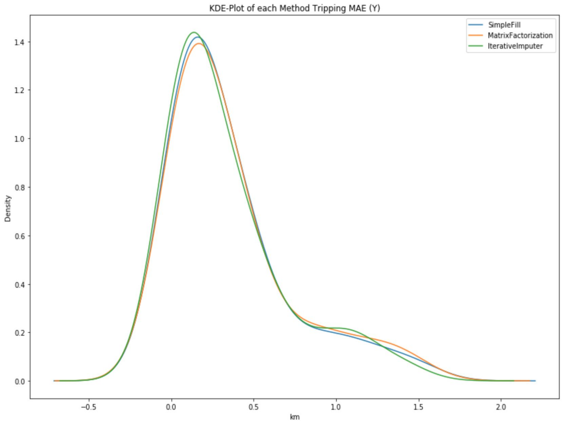
Model	Dwell Horizontal MAE	Trip Horizontal MAE	Dwell Vertical MAE	Trip Vertical MAE
Simple Fill	173.1982 m	346.5374 m	163.0607 m	258.3705 m
Matrix Factorization	127.1270 m	358.3617 m	109.1262 m	271.0766 m
Iterative Imputer	215.5642 m	331.3778 m	173.6039 m	251.7646 m

Results: Dwells vs Trips









Case Study #1
Success

Case Study #2
Failure

Summary

We investigated the application of imputation methods for movement prediction

- The methods can predict missing records with average 225 meters error
- The methods can predict daily trajectories with average 345 meters error
- The performance of the models depends on the diversity of movement behaviour

Thank you!
Questions?

amj301@usask.ca , sem311@usask.ca