UNIVERSITÄT BERN

Predicting Future Locations of Users based on History Data

Dominic Kohler
Institute of Computer Science
Communications and Distributed Systems (CDS)
Universitaet Bern

28.04.2017, CDS Seminar

Outline

UNIVERSITÄT BERN

- Motivation
- Algorithms and Features
  - Difference in Leaving Time
  - Single/Ensemble Predictors
- Evaluation & Results
  - Temporal & Hybrid Features
  - User Movement
- Conclusion

Motivation - Goal

UNIVERSITÄT BERN

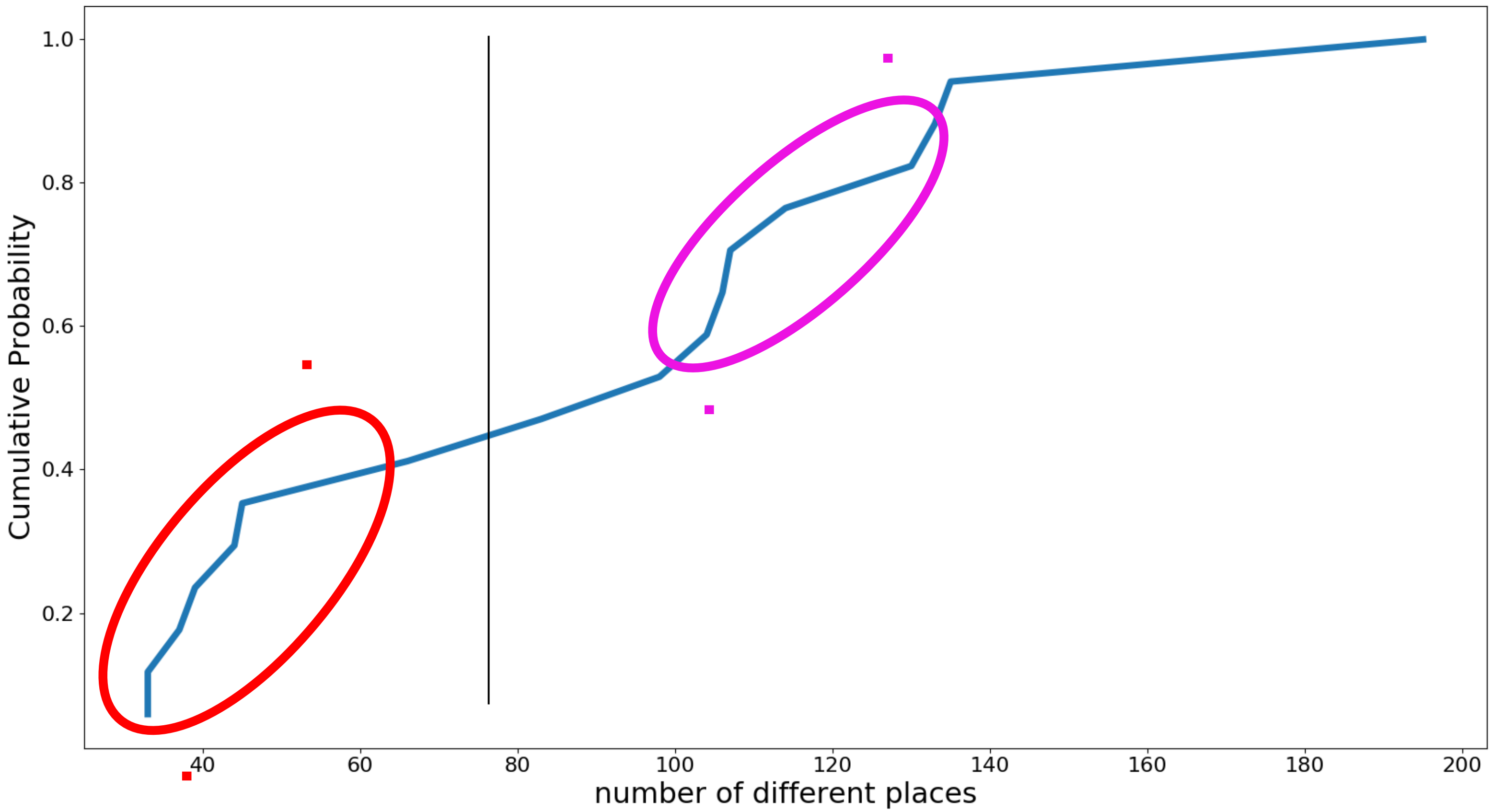
- Improve Location based Services
  - Ubiquity of smartphones „opened new doors"

Predict Future Locations of a User

- Supervised Machinelearning Task
  - Set(User,features) → placeID prediction

Quality of extracted Users

UNIVERSITÄT
BERN



Difference in Leaving Time

UNIVERSITÄT BERN

Place_id	Leaving_time	Next_place_id
1	07.40	2
2	08.30	1
1	07.54	3
3	07.15	5

Table 1: visit_10min

place_id	Leaving_time	Next_place_id
1	07.50	?

Table 2: Instance to classify

$$currentDiff = |\, T_{leavingTime} - I_{leavingTime} \,|$$

$$if\ (currentDiff < smallestDiff\ \&\&\ currentDiff < Threshold\ )$$

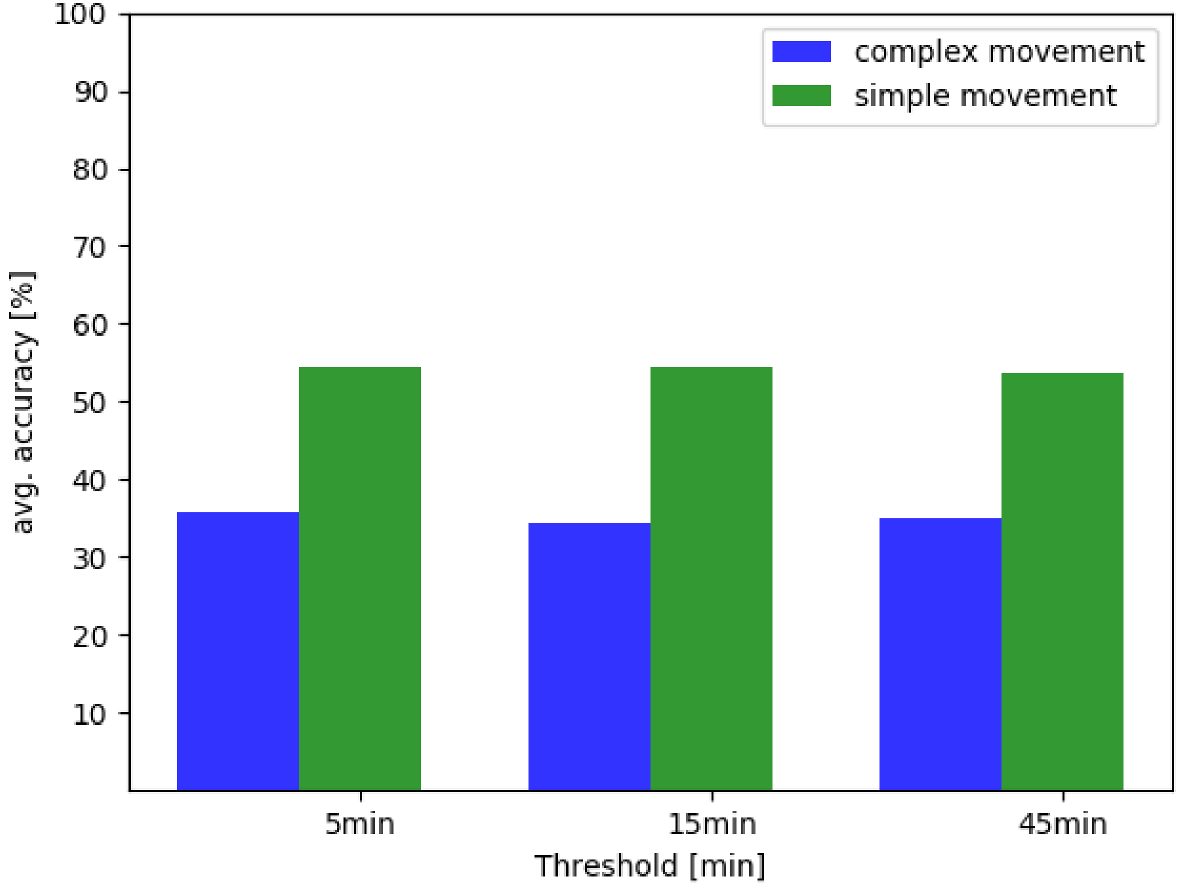
$$I_{nextPlaceID} = T_{nextPlaceId}$$

$$smallestDiff = currentDiff$$

Difference in Leaving Time - Rationale

- Decision Tree if leaf node ambigous → home or office as next_pid
  - Allows for adaption in pattern changes (Holiday,new Home…)

Difference in Leaving Time - Results



J48 + Difference in Leaving Time

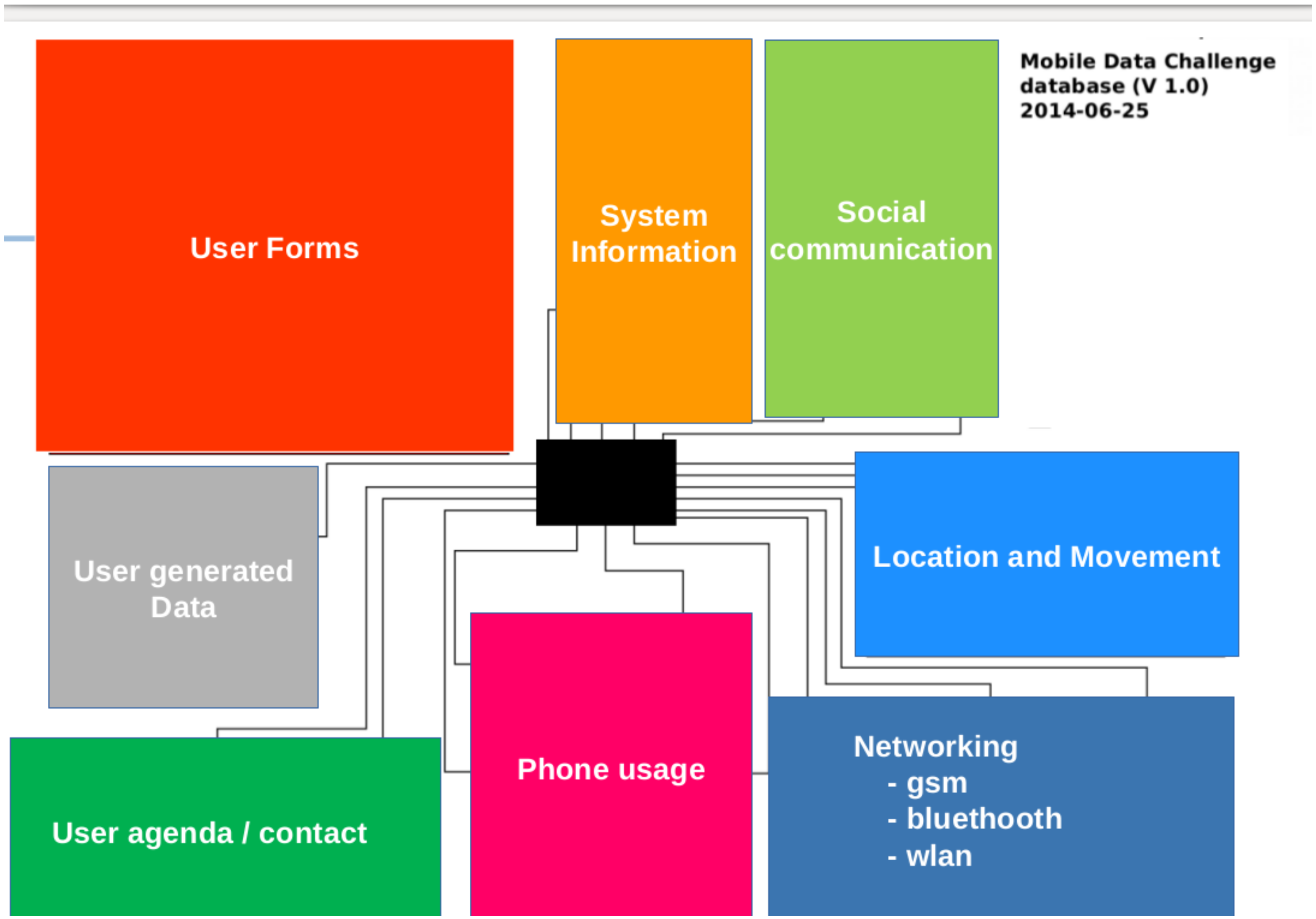
UNIVERSITÄT BERN

- Stacking: J48 and DLT as Base Learner

- Mispredicted Instances from J48 as Input for DLT
  - DLT could not improve overall accuracy

For some users could not even predict one place correctly

MDC Dataset - Categories



Selected Features

UNIVERSITÄT BERN

- Temporal
  - Duration
  - Leavingtime
  - Weekday
- System
  - Charging
  - Profile
- Acceleration
  - Average. Delta
- Networking
  - #WLAN IDs
  - #GSM IDs

Not Selected Features

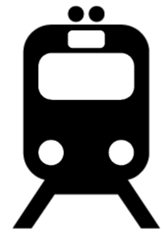
- Application
- Media
- Calllog
- Calendar

General Problems:

- high variance in data
- sparse data

Places - Features

Charging: True
Profile: General
#Wlans: 1<=x<5
#GSM: 6<=x<12



Charging: False
Accel: high
Duration: 0<=x<40

Charging: True
Profile: Silent/Meeting
#Wlans: 5<=x<10
Weekday: MON-FRI
Leavingtime: 09-12.30, 14-18

Decison Tree Model (J48)

Nbr of correct/wrong prediction

```
nbrwlans = 5<=x<10
|   duration = 300<=x<480
|   |   leaving_time = 14:00<=x<18:00: 2 (25.0/1.0)
|   |   leaving_time = 23:00<=x<07:00: 8 (1.0)
|   |   leaving_time = 12:30<=x<14:00: 2 (7.0)
|   |   leaving_time = 18:00<=x<19:00: 2 (0.0)
|   |   leaving_time = 07:00<=x<09:00: 8 (2.0)
|   |   leaving_time = 09:00<=x<12:30: 8 (1.0)
|   |   leaving_time = 19:00<=x<23:00: 2 (0.0)
|   duration = 140<=x<200: 2 (23.0/1.0)
|   duration = 480<=x<2880
|   |   leaving_time = 14:00<=x<18:00: 8 (14.0/6.0)
|   |   leaving_time = 23:00<=x<07:00: 8 (9.0)
|   |   leaving_time = 12:30<=x<14:00: 8 (1.0)
|   |   leaving_time = 18:00<=x<19:00: 2 (3.0)
|   |   leaving_time = 07:00<=x<09:00: 8 (5.0)
|   |   leaving_time = 09:00<=x<12:30: 8 (11.0)
|   |   leaving_time = 19:00<=x<23:00: 2 (3.0)
|   duration = 60<=x<100
|   |   nbrgsms = 0<=x<3: 2 (20.0/1.0)
|   |   nbrgsms = 3<=x<6: 33 (8.0/3.0)
```

Confusion-Matrix of Output Predictions

a	b	c	d	e	f	g	h	i	j	← Classified as
354	0	0	0	0	0	0	0	0	0	a
0	179	0	0	0	0	4	12	0	0	b
0	0	144	0	0	0	0	0	1	0	c
0	0	0	178	0	0	0	0	0	0	d
0	0	0	0	298	0	0	0	0	0	e
0	0	0	0	0	7	0	0	0	0	f

Adding „Profile“ Feature

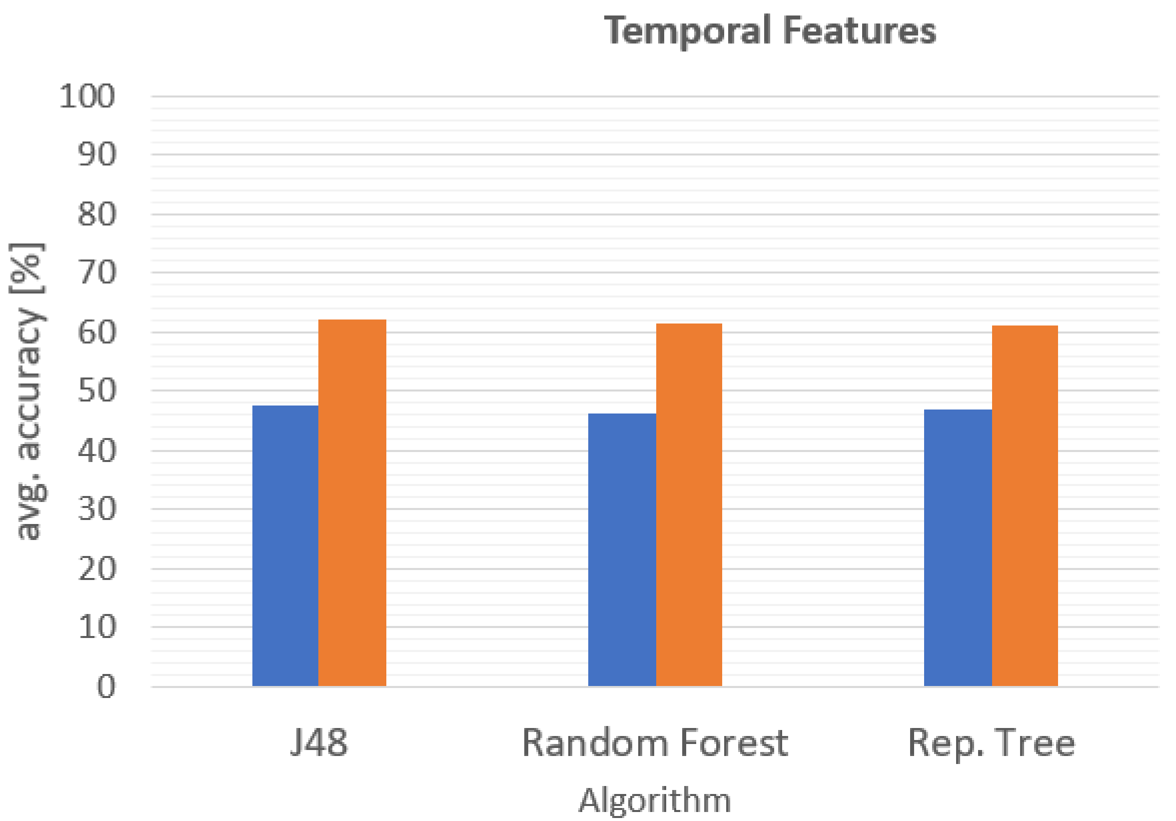
UNIVERSITÄT BERN

```
nbrwlans = 5<=x<10
|   profile = pager: 2 (0.0)
|   profile = outdoor: 2 (0.0)
|   profile = silent
|   |   nbrgsms = 0<=x<3: 2 (212.0/27.0)
|   |   nbrgsms = 3<=x<6: 33 (8.0/2.0)
|   |   nbrgsms = 6<=x<12: 2 (0.0)
|   |   nbrgsms = x>11: 2 (0.0)
|   profile = general
|   |   duration = 300<=x<480: 8 (5.0/1.0)
|   |   duration = 140<=x<200: 8 (1.0)
|   |   duration = 480<=x<2880: 8 (33.0)
|   |   duration = 60<=x<100: 2 (4.0/3.0)
|   |   duration = 200<=x<300: 8 (3.0)
|   |   duration = 0<=x<40: 22 (95.0/54.0)
|   |   duration = 100<=x<140: 2 (4.0/2.0)
|   |   duration = 40<=x<60
|   |   |   leaving_time = 14:00<=x<18:00: 2 (7.0/3.0)
|   |   |   leaving_time = 23:00<=x<07:00: 64 (1.0)
|   |   |   leaving_time = 12:30<=x<14:00: 7 (2.0/1.0)
|   |   |   leaving_time = 18:00<=x<19:00: 2 (0.0)
|   |   |   leaving_time = 07:00<=x<09:00: 2 (0.0)
|   |   |   leaving_time = 09:00<=x<12:30: 2 (0.0)
|   |   |   leaving_time = 19:00<=x<23:00: 2 (0.0)
```

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	← Classified as
353	0	0	0	0	0	0	0	0	0	a
0	194	0	0	0	0	0	1	0	0	b
0	0	141	0	0	0	0	0	1	0	c
0	0	0	178	0	0	0	0	0	0	d
0	0	0	0	298	0	0	0	0	0	e
0	0	0	0	0	7	0	0	0	0	f

Evaluation - Single Predictors



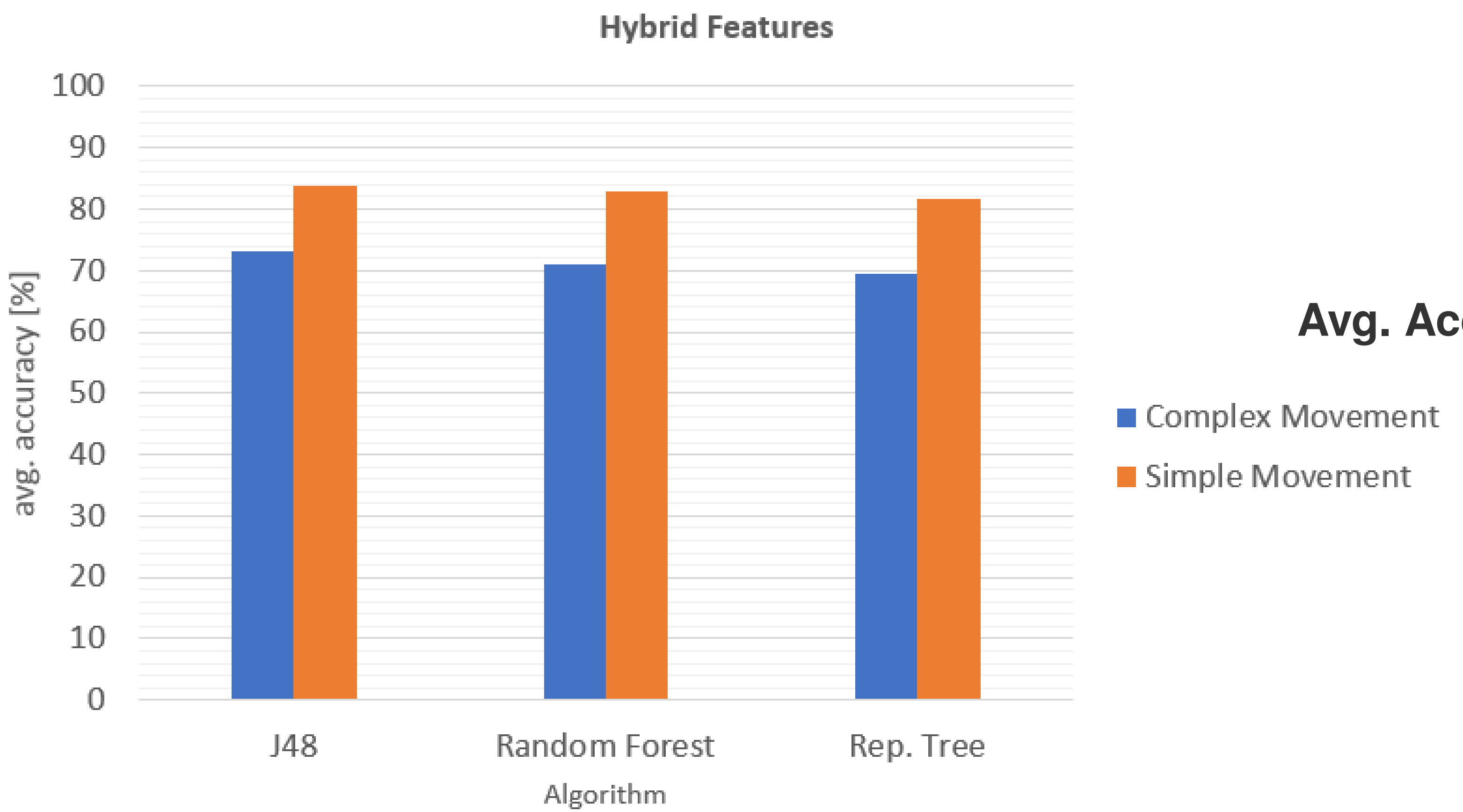
Avg. Accuracy

Complex Movement **47%**

Simple Movement **62%**

10 –Folds Crossvalidation

Evaluation - Single Predictors



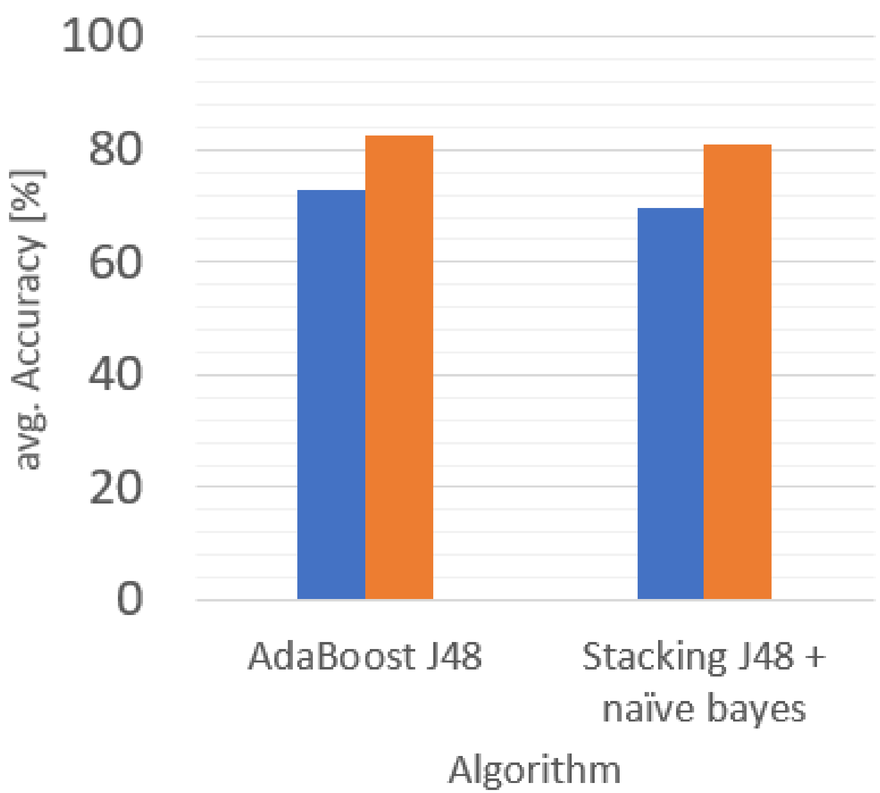
Avg. Accuracy

72%

84%

10 –Folds Crossvalidation

Evaluation - Ensemble Predictors



Avg. Accuracy

Complex Movement	**72%**
Simple Movement	**81%**

10 –Folds Crossvalidation

Related Work

- K. Aberer et. Al, „Next Place Prediction usin Mobile Data“
  - Users:
    - Simple Movement Pattern: avg. Accuracy: 71.85%
    - Heterogenous Movement: avg. Accuracy: 54.27%

—Temporal Features
—Holiday Detection
—Difference in Leavingtime

Conclusion

- Hybrid features improve accuracy
  -WLAN and GSM are potent features

- Prediction accuracy depends on user quality

- Ensemble methods did not significantly improve accuracy