Chonnam National University
Advanced Network Lab

Modeling and Discovering Human Behavior from Smartphone Sensing Life-Log Data

Rischan Mafrur

June 1, 2015

Supervised by Professor Deokjai Choi

Outlines

- **Introduction**
  - Background and Problem Statements
- **Methods**
  - Data Collection
  - Data Preprocessing and Features Extraction
  - Our approach details
- **Goal**
  - Discovering user behavior based on their smartphone life-log
  - Modeling human behavior, evaluated for user identification

Thesis Background and Problem Statements

Common approaches

- Most of previous works is focus on one sensor for one purpose.

Realistic Dataset

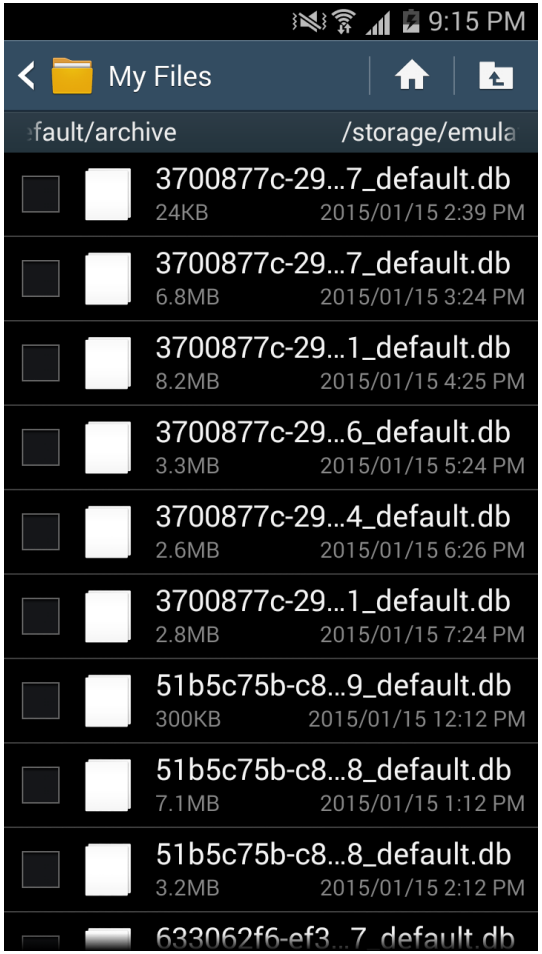
- User has different types and brand of smartphone.
- We could not expect the human actions and activities.
- There is no ideal data collection which running for 24 hour non-stop.
- There is no ideal data collection that can record all of data without any data loss.

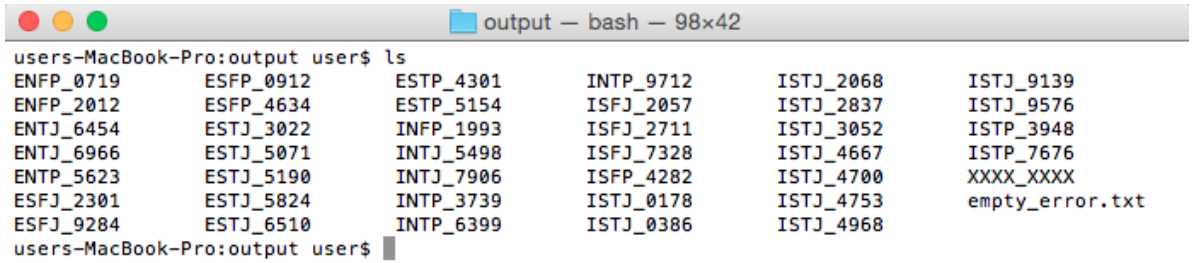
We decide to use many of sensors rather than focus only one sensor. Our proposed method tried to deal with those situations.

Data Collection

1. To collect user personal data from their smartphone, we developed an Android application data collector based on *Funf* library.
2. This application follows opportunistic sensing method.
3. We asked 47 students to use our application around less than 2 months.
4. The size of all of the data is around **28.7 GB**.
5. In this research, We used only 37 students data.
6. The raw data looks like can be seen in the next slide.

Raw Data looks like





_id	name	timestamp	value
Filter	Filter	Filter	Filter
1	HardwareInfoProbe	1404116791.153	{"androidId":"df03d9eae60fecb3","bluetoothMac":"88:9...
2	SmsProbe	1403883215.0	{"address":"15444302","body":"{\"ONE_WAY_HASH\":\"...
3	SmsProbe	1403883087.0	{"address":"15444302","body":"{\"ONE_WAY_HASH\":\"...
4	CallLogProbe	1404112293.842	{"_id":4108,"date":1404112293842,"duration":29,"name":...
5	SmsProbe	1403772474.744	{"address":"01066296348","body":"{\"ONE_WAY_HASH\...
6	SmsProbe	1403697660.0	{"address":"0220338500","body":"{\"ONE_WAY_HASH\"...
7	CallLogProbe	1404111655.487	{"_id":4107,"date":1404111655487,"duration":75,"name":...
8	SmsProbe	1403697242.0	{"address":"0625101111","body":"{\"ONE_WAY_HASH\"...
9	CallLogProbe	1404111605.734	{"_id":4106,"date":1404111605734,"duration":0,"name":"...
10	SmsProbe	1403609220.0	{"address":"0622395000","body":"{\"ONE_WAY_HASH\"...
11	SmsProbe	1403540566.46	{"address":"01029738808","body":"{\"ONE_WAY_HASH\...
12	CallLogProbe	1404111295.005	{"_id":4105,"date":1404111295005,"duration":0,"name":"...
13	SmsProbe	1403540204.981	{"address":"01086198051","body":"{\"ONE_WAY_HASH\...

Human behavior in case of smartphone sensing

- What is the human behavior in case of smartphone sensing?.
  - Human daily activities which carried out continuously
- In terms of human daily activities, we have to consider about four things:
  - What kind of activity (e.g meeting, studying, exercising)
  - When (e.g around 9 AM)
  - Location (e.g Lab)
  - Human Interaction (e.g all lab's members)

What, When, Where, Who/with Whom

Proposed Sensors Description

Proposed Sensors data

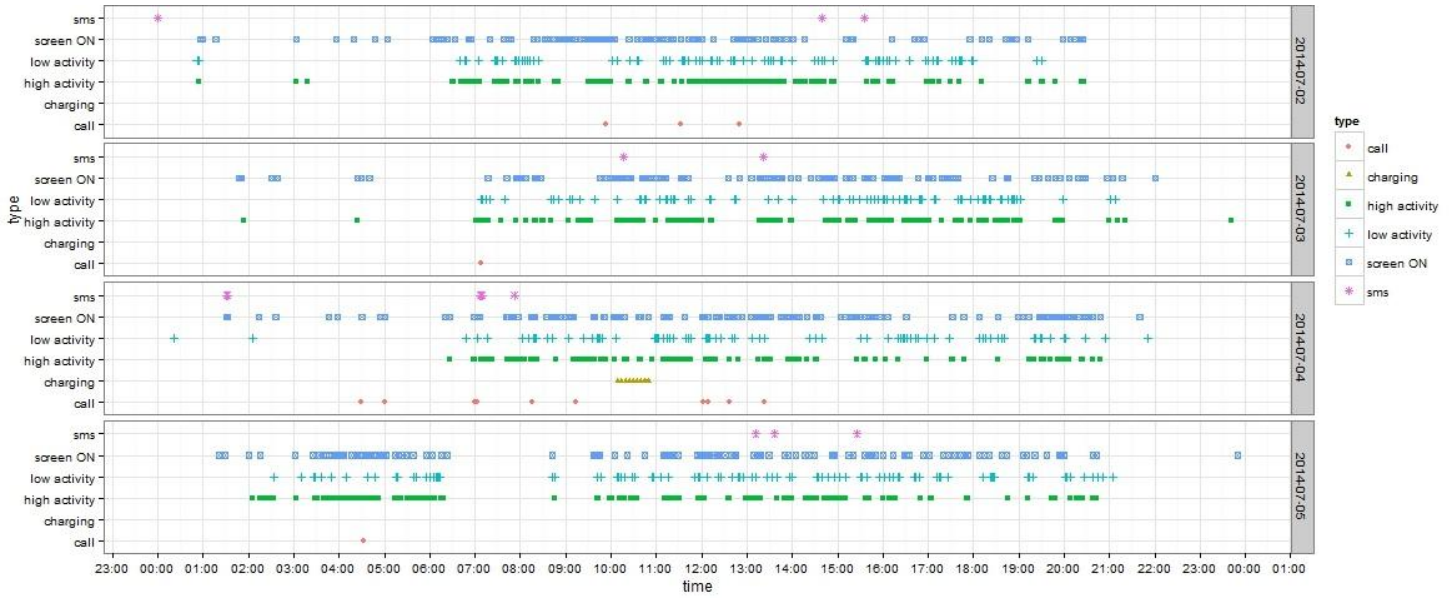
(every single data has timestamp)

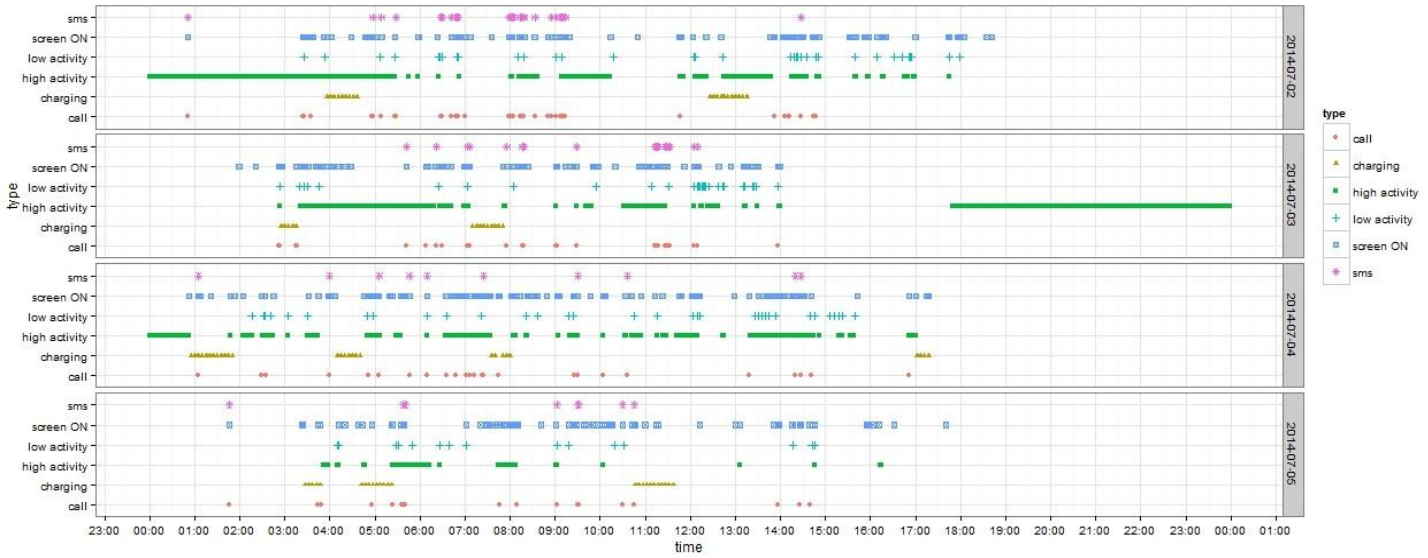
- What kind of Human Activity
  - Activity [none, low, high]
- Human Location
  - GPS [longitude, latitude]
  - Nearby Bluetooth [list of nearby Bluetooth]
  - Nearby Wi-Fi [lists of nearby AP]
- Human Interaction (Human->Human)
  - Call [incoming, outgoing, missed]
  - SMS [sent, received]
- Human Interaction (Human -> Smartphone)
  - Battery [time charging]
  - Run apps [name of apps]
  - Screen [screen ON, screen OFF]

Table 2-2. List of probes and types

No.	Name of Probes	Explanation	Used
On Request Data			
1.	SimpleLocationProbe	GPS data (user location)	X
2.	WifiProbe	Nearby Wi-Fi signals	X
3.	BluetoothProbe	Nearby Bluetooth signals	X
4.	BatteryProbe	Battery status	X
Historical Data			
1.	CallLogProbe	User call log	X
2.	SmsProbe	User SMS log	X
3.	ApplicationsProbe	List of application installed	
4.	HardwareInfoProbe	User's smartphone hardware info	
5.	BrowserBookmarksProbe	User Bookmarks	
6.	BrowserSearchesProbe	User Browser log	
7.	ContactProbe	User contact (phonebook)	
Continuous Data			
1.	LightSensorProbe	Measures the ambient light level (illumination) in lx	
2.	ProximitySensorProbe	Measures the proximity of an object in cm relative to the view screen of a device.	
3.	TemperatureSensorProbe	Measures the temperature of the device in degrees Celsius (°C).	
4.	MagneticFieldSensorProbe	Measures the ambient geomagnetic field (x, y, z) in µT	
5.	PressureSensorProbe	Measures the ambient air pressure in hPa or mbar.	
6.	ScreenProbe	Screen phone (on and off)	X
7.	RunningApplicationsProbe	List of running applications	X
8.	ActivityProbe	User activity log based on accelerometer sensor (none, low, and high activity)	X

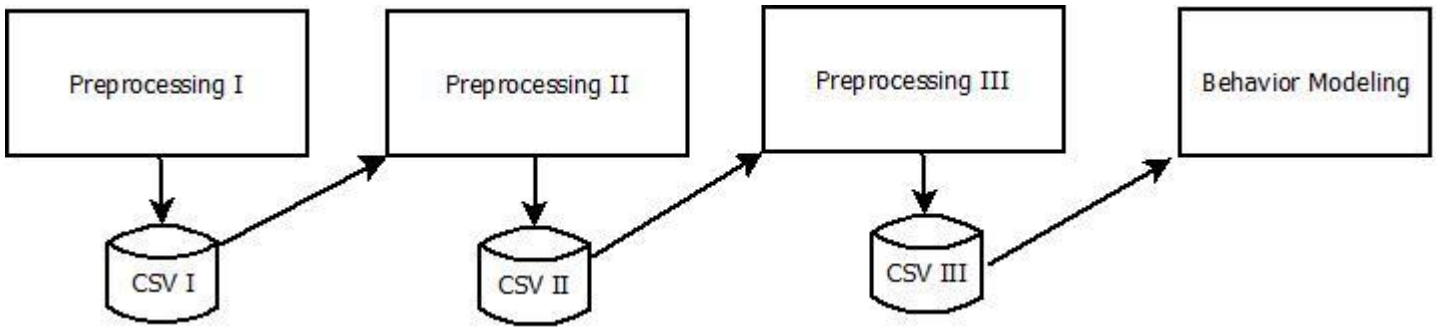
Example data plot from two students





Technical Explanation

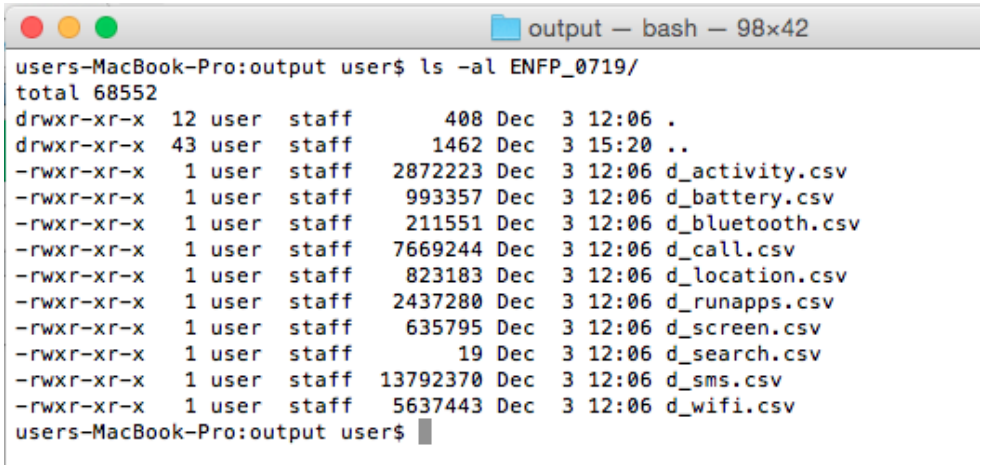
Pre-Processing Summarization



- Preprocessing I
  - Data cleansing
  - Removing duplication and noisy data
  - Select the most important data
- Preprocessing II
  - Features Extraction applied in here
- Preprocessing III
  - Features Aggregation
  - Adding new values based on features
  - Fitting the features data before modeling behavior applied

Source code can be access on ***http://github.com/rischanlab/Rfunf***

Preprocessing I



The function of this module are:

- Removing duplication
- Data cleansing
- Select the most important data

Preprocessing I output data looks like

ENFP_0719 — more — 98×42

```
"","when","latitude","longitude"
"1","2014-06-30 08:26:36","35.1754255","126.9130674"
"2","2014-06-30 08:31:30","35.1754281","126.9130184"
"3","2014-06-30 08:36:30","35.1754225","126.9130536"
"4","2014-06-30 08:41:30","35.1753982","126.9130597"
"5","2014-06-30 08:46:30","35.1753991","126.9130549"
"6","2014-06-30 08:51:30","35.1753997","126.9130598"
"7","2014-06-30 08:56:31","35.1754333","126.9130484"
"8","2014-06-30 09:01:30","35.1754143","126.9130638"
"9","2014-06-30 09:06:30","35.1754126","126.9130687"
"10","2014-06-30 09:11:30","35.1754142","126.9130553"
"11","2014-06-30 09:16:30","35.1753987","126.913074"
"12","2014-06-30 09:21:30","35.1754178","126.9130686"
"13","2014-06-30 09:26:30","35.1754176","126.9130779"
"14","2014-06-30 09:31:30","35.1753914","126.9130496"
"15","2014-06-30 09:36:30","35.1754066","126.9130196"
"16","2014-06-30 09:41:30","35.1753833","126.9130507"
"17","2014-06-30 09:46:30","35.1753883","126.9130474"
"18","2014-06-30 09:51:52","35.174465","126.9135807"
"19","2014-06-30 09:56:30","35.174465","126.9135807"
"20","2014-06-30 10:01:30","35.1735813","126.9119189"
"21","2014-06-30 10:06:53","35.1744781","126.9137426"
"22","2014-06-30 10:11:30","35.1735813","126.9119189"
"23","2014-06-30 10:16:52","35.174502","126.9137499"
"24","2014-06-30 10:21:30","35.1749091","126.9124954"
"25","2014-06-30 10:26:30","35.1749091","126.9124954"
"26","2014-06-30 10:31:30","35.1749091","126.9124954"
```

ENFP_0719 — more — 98×42

```
"","timestamp","duration","package"
"1","2014-06-30 08:26:31","12.401","edu.mit.media.funf.wifiscanner"
"2","2014-06-30 08:26:43","2.055","com.nhn.android.search"
"3","2014-06-30 08:26:45","9.183","com.buzzpia.aqua.launcher"
"4","2014-06-30 08:26:54","15.32","edu.mit.media.funf.wifiscanner"
"5","2014-06-30 08:27:10","38.126","com.rechild.advancedtaskkiller"
"6","2014-06-30 08:27:48","3.015","edu.mit.media.funf.wifiscanner"
"7","2014-06-30 08:27:51","6.015","com.buzzpia.aqua.launcher"
"8","2014-06-30 08:27:57","2.005","edu.mit.media.funf.wifiscanner"
"9","2014-06-30 08:27:59","2.006","com.buzzpia.aqua.launcher"
"10","2014-06-30 08:28:01","2.044","com.kakao.talk"
"11","2014-06-30 08:28:03","6.042","com.buzzpia.aqua.launcher"
"12","2014-06-30 08:28:09","63.362","com.nhn.android.search"
"13","2014-06-30 08:29:12","0.841","com.buzzpia.aqua.launcher"
"14","2014-06-30 08:30:02","55.247","com.buzzpia.aqua.launcher"
"15","2014-06-30 08:33:35","9.898","com.buzzpia.aqua.launcher"
"16","2014-06-30 08:33:46","4.015","com.buzzpia.aqua.launcher"
"17","2014-06-30 08:33:50","68.308","com.kakao.talk"
"18","2014-06-30 08:34:58","9.017","com.kakao.talk"
"19","2014-06-30 08:35:07","3.006","com.buzzpia.aqua.launcher"
"20","2014-06-30 08:35:10","205.598","com.facebook.katana"
"21","2014-06-30 08:38:36","1.245","com.buzzpia.aqua.launcher"
"22","2014-06-30 08:40:27","5.011","com.buzzpia.aqua.launcher"
"23","2014-06-30 08:40:32","61.606","com.joeware.android.gpulumera"
"24","2014-06-30 08:41:34","13.806","com.buzzpia.aqua.launcher"
```

Pre-processing II (Features Extraction)

List of Sensors Data and Features Values.

No	Name of Probes	Value1	Value2	Value3
1.	ActivityProbe	Status (“*none*”,”*low*”, and ”*high*”)		
2.	SimpleLocationProbe	Latitude	Longitude	
3.	WifiProbe	List of nearby SSID	MAC	Signal strength (dB)
4.	BluetoothProbe	List of nearby Bluetooth devices		
5.	BatteryProbe	Status (“*discharging*”,”*full*”, and ”*charging*”)		
6.	ScreenProbe	ON/OFF		
7.	RunningApplicationsProbe	Apps name	Duration	
8.	CallLogProbe	Number	Types	Duration
9.	SmsProbe	Number	Types	Text length

Preprocessing II output data looks like

output2 — more — 98×42

```
"","time","type","value1","value2","value3"
"5135","2014-07-01 00:01:55","activity","none","",""
"5136","2014-07-01 00:01:56","activity","none","",""
"96154","2014-07-01 00:02:54","battery","full","",""
"166061","2014-07-01 00:02:54","location","37.53724098","126.96960174",""
"487488","2014-07-01 00:02:58","wifi","2-607","44:ed:57:01:f9:ac","-55"
"5137","2014-07-01 00:03:56","activity","none","",""
"5138","2014-07-01 00:03:57","activity","none","",""
"5139","2014-07-01 00:05:56","activity","none","",""
"96155","2014-07-01 00:07:55","battery","full","",""
"166062","2014-07-01 00:07:55","location","37.53724098","126.96960174",""
"5140","2014-07-01 00:07:56","activity","none","",""
"5141","2014-07-01 00:07:57","activity","none","",""
"487489","2014-07-01 00:07:58","wifi","2-607","44:ed:57:01:f9:ac","-54"
"487490","2014-07-01 00:07:58","wifi","2-507","44:ed:57:01:55:ca","-86"
"5142","2014-07-01 00:09:55","activity","none","",""
"5143","2014-07-01 00:09:56","activity","none","",""
"5144","2014-07-01 00:11:55","activity","none","",""
"5145","2014-07-01 00:11:56","activity","none","",""
"96156","2014-07-01 00:12:55","battery","full","",""
"166063","2014-07-01 00:12:55","location","37.53724098","126.96960174",""
"487491","2014-07-01 00:12:58","wifi","2-607","44:ed:57:01:f9:ac","-61"
"487492","2014-07-01 00:12:58","wifi","2-507","44:ed:57:01:55:ca","-87"
"487493","2014-07-01 00:12:58","wifi","sparrow","64:e5:99:c8:06:d2","-92"
"5146","2014-07-01 00:13:57","activity","none","",""
"5147","2014-07-01 00:13:58","activity","none","",""
"5148","2014-07-01 00:15:56","activity","none","",""
"5149","2014-07-01 00:15:57","activity","none","",""
"96157","2014-07-01 00:17:55","battery","full","",""
"166064","2014-07-01 00:17:55","location","37.53724098","126.96960174",""
"5150","2014-07-01 00:17:56","activity","none","",""
"5151","2014-07-01 00:17:57","activity","none","",""
"487494","2014-07-01 00:17:58","wifi","2-607","44:ed:57:01:f9:ac","-54"
"487495","2014-07-01 00:17:58","wifi","2-507","44:ed:57:01:55:ca","-87"
"5152","2014-07-01 00:19:55","activity","none","",""
"5153","2014-07-01 00:19:57","activity","none","",""
"5154","2014-07-01 00:21:55","activity","none","",""
"5155","2014-07-01 00:21:56","activity","none","",""
"96158","2014-07-01 00:22:55","battery","full","",""
"166065","2014-07-01 00:22:55","location","37.53724098","126.96960174",""
"487496","2014-07-01 00:22:58","wifi","2-607","44:ed:57:01:f9:ac","-52"
ENFP_2012.csv
```

Preprocessing III

- Machine Time -> Human Time (round time values)
  - < :30 -> round down
  - > :30 -> round up
- Change the Location value to *(“same”,”little”,”long”)*
  - 0.0001 degree = 11.1132 m (“little”: “between 0.001 ~ 0.005”)

Discovering Human Behavior (Finding Similar Patterns)

Day1 Week1	Day2 Week1	Day3 Week1	Day4 Week1	Day5 Week1	Day6 Week1	~~Day7 Week1~~	Day1 Week2

Time	Sensor Name	Sensor Value
13:00	location	same
13:00	wifi	1-AP, iptime
14:00	runapps	kakao
14:00	location	long
15:00	runapps	kakao
15:00	location	little

Time	Sensor Name	Sensor Value
13:00	location	same
13:00	wifi	1-AP, iptime
14:00	battery	charging
14:00	wifi	D-link
15:00	runapps	kakao
15:00	location	little

Group-1 = 13:00,location,same | 13:00,wifi,1-AP,iptime

Group-1 = 13:00,location,same | 13:00,wifi,1-AP,iptime

Group-2 = 15:00,runapps,kakao | 15:00, location, little

Group-2 = 15:00,runapps,kakao | 15:00, location, little

Algorithm (Similarity Detection)

```
Data : D, w
Result : All Detected Group in a Window
grpAll, grpTemp, grpPrevious<- NULL
dataValue, dataValueNext <- NULL
while (D in w) for all of D do
    dataValue <- D.current.day
    dataValueNext <- D.next.day
    grpTemp <- findingSimilarPatterns(dataValue, dataValueNext)

    if (grpTemp in grpPrevious)then
         grpNew <- merge(grpPrevious, grpTemp)
         grpAll <- add(grpNew)
    else
         grpAll <- add(grpTemp)
```

Behavior Profiling/Modeling

We collect all of intersection data between Groups, and mark those data as the user behaviors.

Grouping Result

```
G2,"19:00,location,same|19:00,bluetooth,DTVBluetooth|"
G2,"19:00,location,same|19:00,bluetooth,DTVBluetooth|"
G2,"19:00,location,same|19:00,bluetooth,DTVBluetooth|"
G2,"19:00,location,same|19:00,bluetooth,DTVBluetooth|"
G2,"19:00,location,same|19:00,bluetooth,DTVBluetooth|"
G3,"20:00,location,same|20:00,bluetooth,DTVBluetooth|"
G3,"20:00,location,same|20:00,bluetooth,DTVBluetooth|"
G3,"20:00,location,same|20:00,bluetooth,DTVBluetooth|"
G3,"20:00,location,same|20:00,bluetooth,DTVBluetooth|"
G7,"06:00,screen,OFF|06:00,location,same|"
G7,"06:00,screen,OFF|06:00,location,same|"
G7,"06:00,screen,OFF|06:00,location,same|"
G7,"06:00,screen,OFF|06:00,location,same|"
G7,"06:00,screen,OFF|06:00,location,same|"
G7,"06:00,screen,OFF|06:00,location,same|"
G7,"06:00,screen,OFF|06:00,location,same|"
G7,"06:00,screen,OFF|06:00,location,same|"
G7,"06:00,screen,OFF|06:00,location,same|"
G7,"06:00,screen,OFF|06:00,location,same|"
G7,"06:00,screen,OFF|06:00,location,same|"
G7,"06:00,screen,OFF|06:00,location,same|"
G7,"06:00,screen,OFF|06:00,location,same|"
G9,"12:00,battery,charging|12:00,bluetooth,ESVH-PC|12:00,location,same|12:00,runapps,com.lge.launcher2|"
G9,"12:00,battery,charging|12:00,bluetooth,ESVH-PC|12:00,location,same|12:00,runapps,com.lge.launcher2|"
G9,"12:00,battery,charging|12:00,bluetooth,ESVH-PC|12:00,location,same|12:00,runapps,com.lge.launcher2|"
G9,"12:00,battery,charging|12:00,bluetooth,ESVH-PC|12:00,location,same|12:00,runapps,com.lge.launcher2|"
G9,"12:00,battery,charging|12:00,bluetooth,ESVH-PC|12:00,location,same|12:00,runapps,com.lge.launcher2|"
G9,"12:00,battery,charging|12:00,bluetooth,ESVH-PC|12:00,location,same|12:00,runapps,com.lge.launcher2|"
G9,"12:00,battery,charging|12:00,bluetooth,ESVH-PC|12:00,location,same|12:00,runapps,com.lge.launcher2|"
INTJ_8928_groups.csv
```

Testing for Human Identification

Performance Evaluation

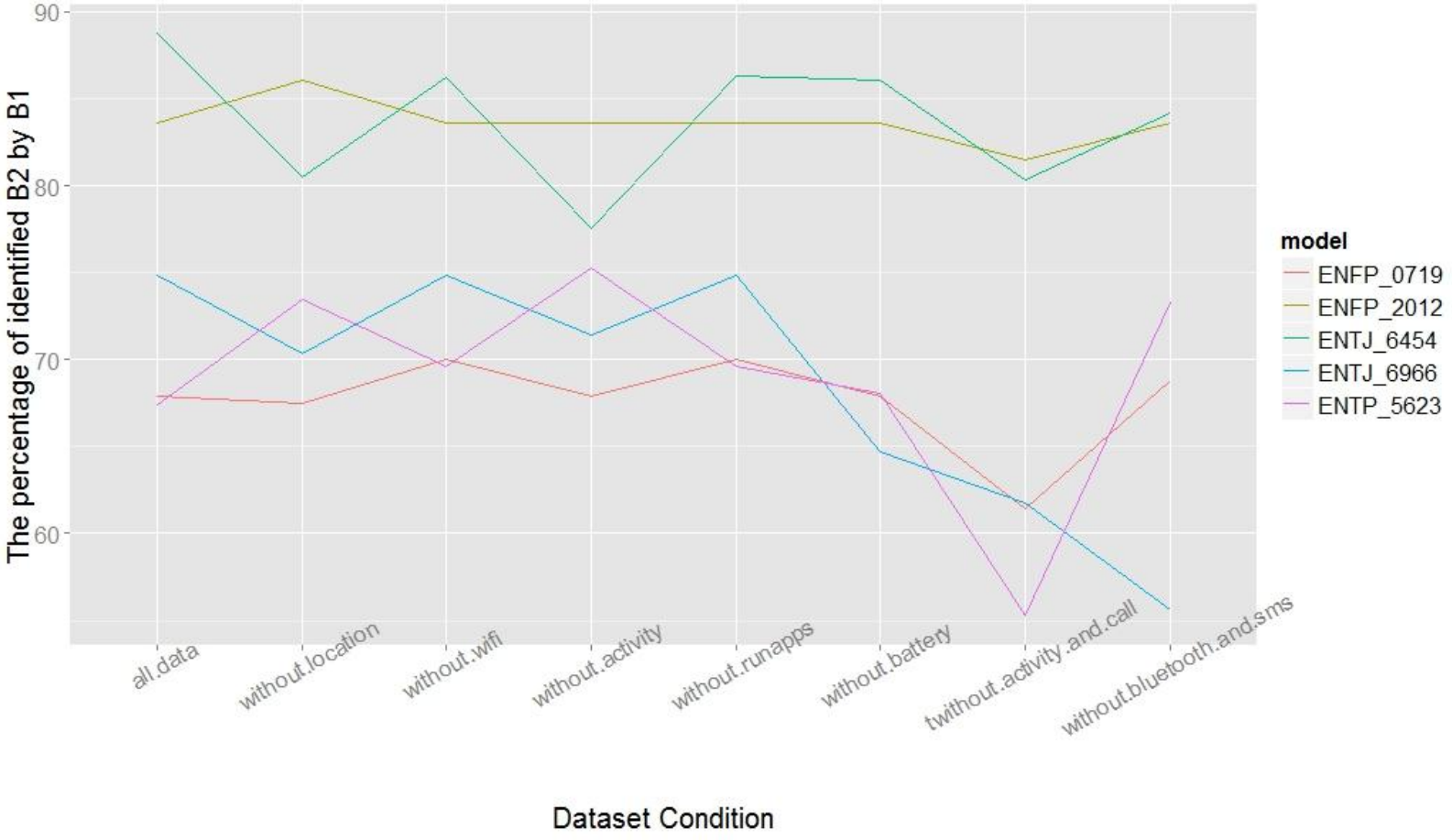
- The total dataset collected around 1 month 20 days
- We Divide all dataset to two parts
  - First month for creating model (first dataset)
  - Remaining dataset for testing performance (second dataset)
- Modeling user behavior based on first dataset,
  - B1: Behavior model/profile.
- Extract and Process the second dataset.
- Apply similarity detection to second dataset with same setting.
  - B2: Set of behavior groups from second dataset.
- Is all of B2 identified by (B1)?.
- How many set of group activities (B2) which identified by (B1), Calculate the percentage of data identified.
- Show result in Table.

Identification result (Only 6 students)

	TEST						
MODEL		ENFP_0719	ENFP_2012	INTJ_5498	ISTJ_3052	ESTJ_5190	ESFP_4634
	ENFP_0719	**67.922**	0	0.4	2.187	0	1.943
	ENFP_2012	0	**83.582**	0	0	0	0
	INTJ_5498	2.178	0	**75.977**	2.087	0	3.401
	ISTJ_3052	2.289	0	0.4	**93.439**	8.232	1.943
	ESTJ_5190	0	0	0	0.099	**22.866**	0
	ESFP_4634	2.289	0	0.977	2.087	0	**89.686**

Full Table can be seen on Appendix, page : 42.

Testing Performance by Removing Some Data Sensors



Full Figure can be seen on Appendix, page : 48.

Conclusions

- In this thesis, we have proposed a approach for daily behavioral pattern mining or discovering human behavior from multiple information (data sensors).
- We use and combine many sensors instead only focus on one sensors because we realize about realistic dataset.
- To evaluate our model, we use our model for human identification.
- Based on identification result, we can see that our model is good enough for user identification. We have tried to remove one or more data sensors and accuracy values is still good enough.

Future Works

- Change the size of window (find optimal window), current size is static (2 days).
- Use different time precision (10, 15, 30 minutes), current is one hour.
- Use window in vertical when compare between days (It will compare between same days in different week).
- Update Model (It will make system better and better but how about storage).
- Find best number of days for creating model.
- Big goal of this research is to develop smart personal assistant.