Utilizing In-Store Sensors for Revisit Prediction

Sundong Kim (sundong.kim@kaist.ac.kr) / Jae-Gil Lee @ KAIST

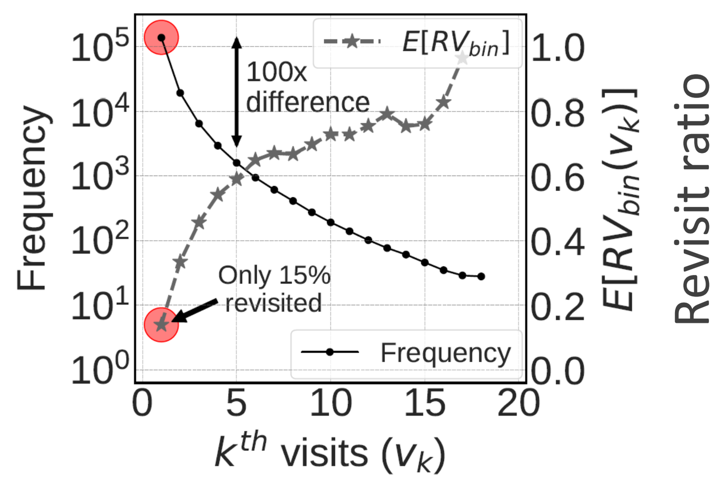
For details!

CEO: Hi Sundong, Welcome to join our team. As you already know, many customers are looking for our stores, but our revenue does not seem to increase. Do you have any thought?

Hi John, I'm glad to part of this family. I heard that you have been tracking customer mobility patterns, let's have a look at the data.

CEO: According to the managers, our main revenue comes from the VIP customers. How can we increase the number of VIP customers?

Unfortunately, the data shows that majority of people visited our store just once. We were losing potential customers! Which customers make purchases at our stores?

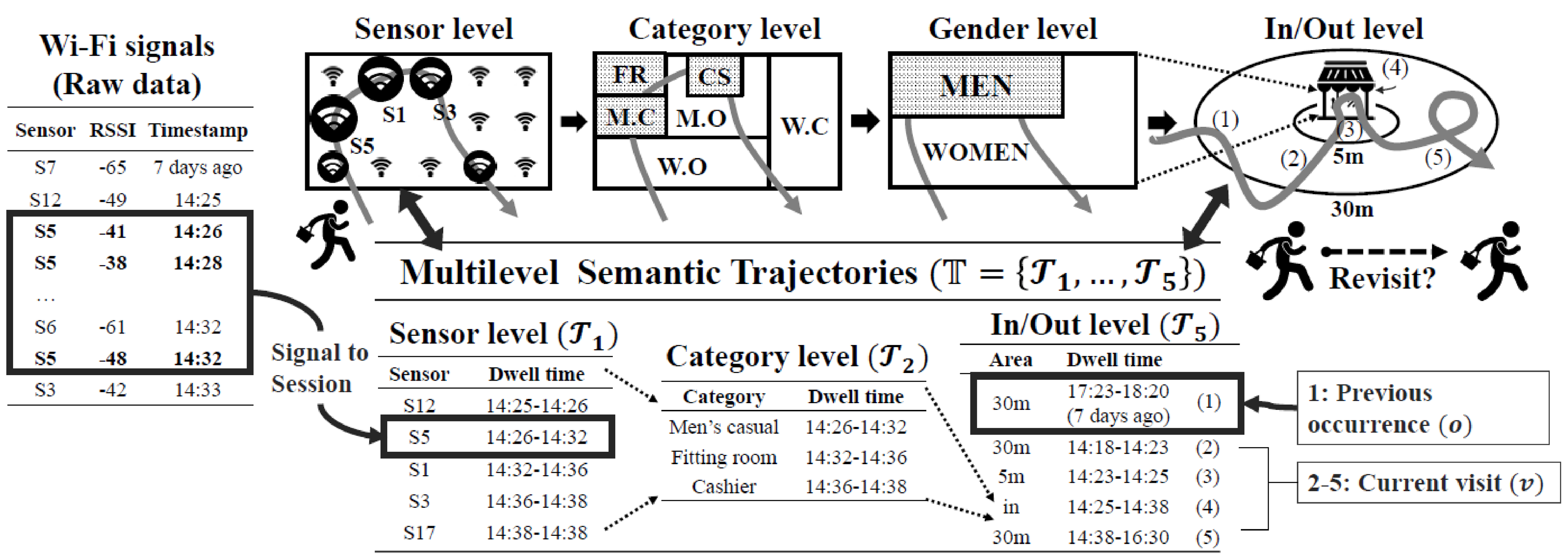


Well, how about applying a marketing strategy that encourages customers to revisit the store? I think it would be great if we are able to figure out customer revisit intention first.

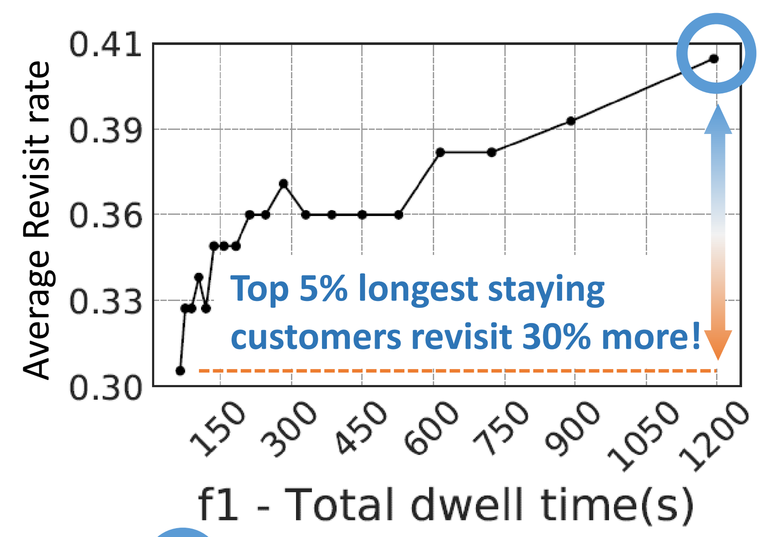
CEO: It would be amazing if we can predict customer revisit from their movements! Let's work on that direction.

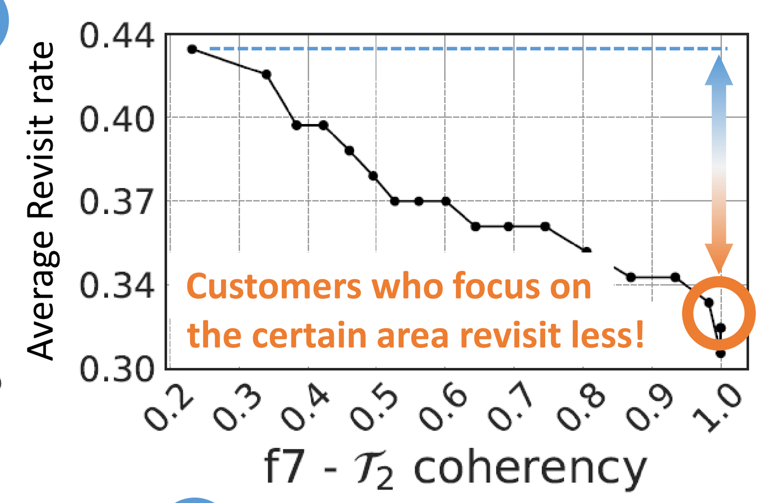
- Note that this dialogue is written for easy understanding.

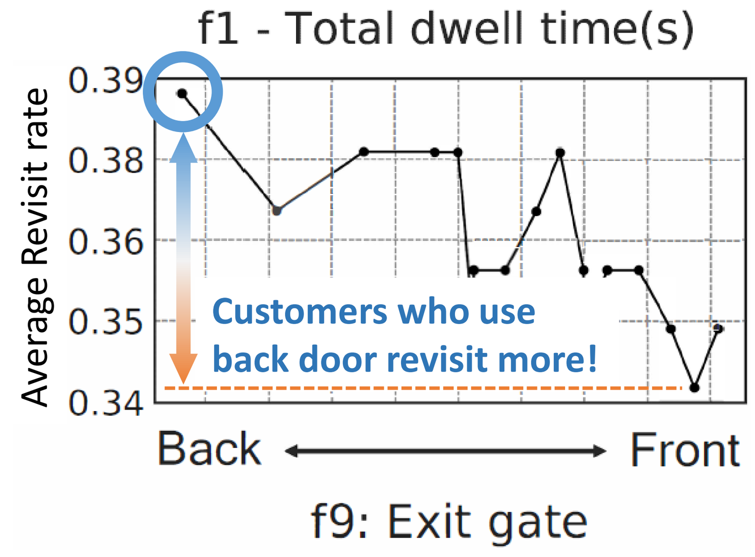
Using Wi-Fi signals, we derive multi-level trajectories to analyze customer mobility patterns for each semantic level. We found out that weak signals, which may be regarded as noise, are important ad-hoc information when no user address is given.

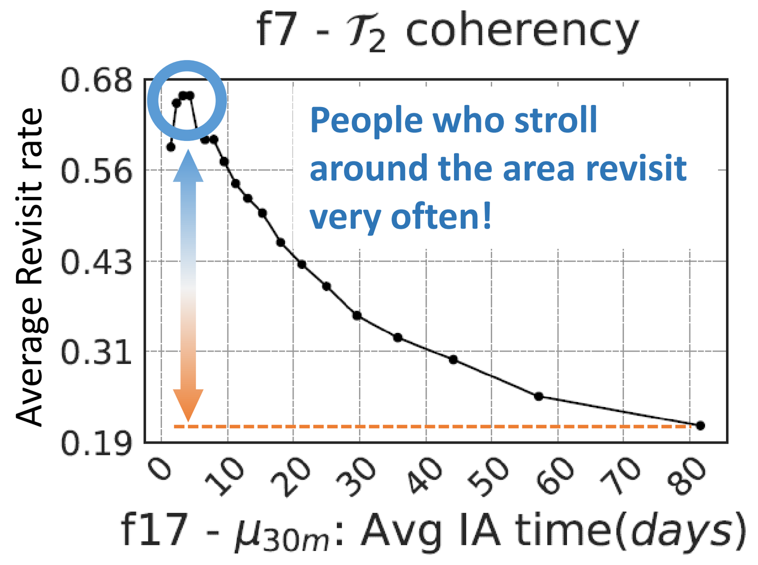


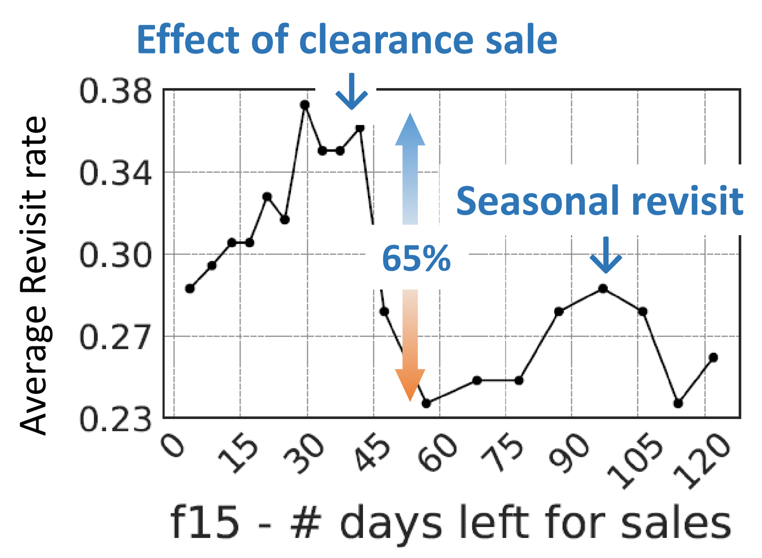
We carefully designed handcrafted features from their mobility patterns. Below figures introduce interesting correlations between feature values and revisit ratio. ☺



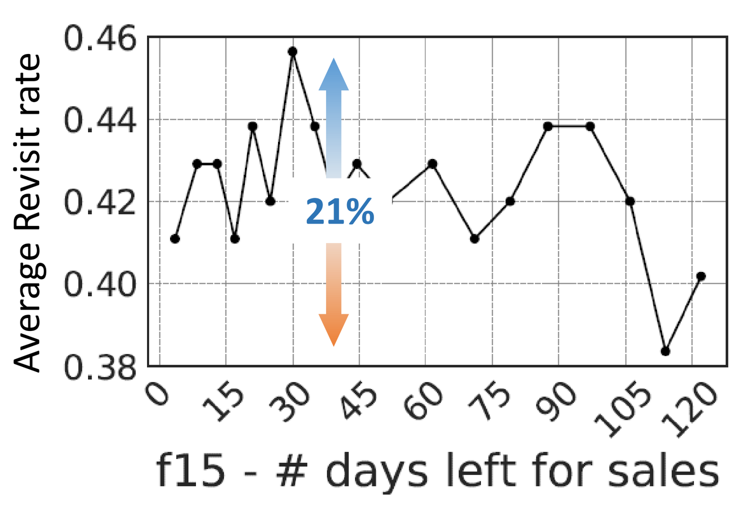








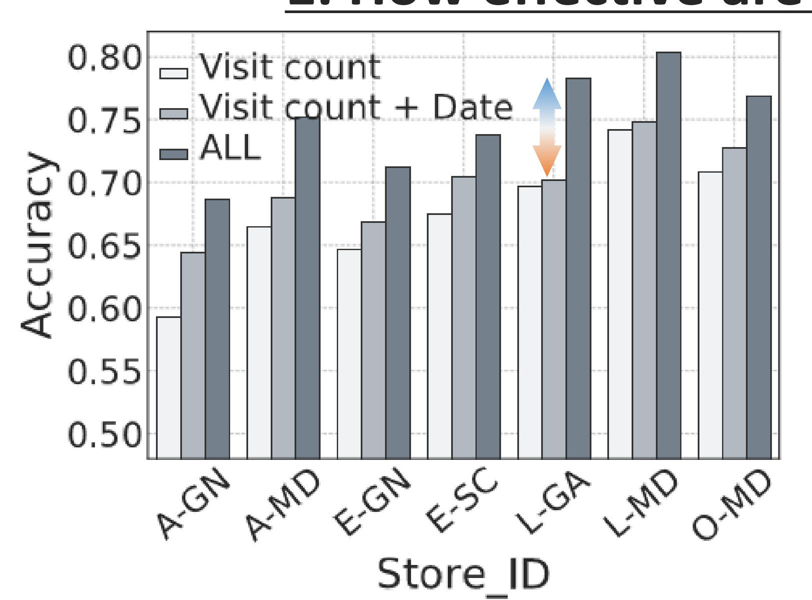
(a) First-time visitors: Prone to special events.



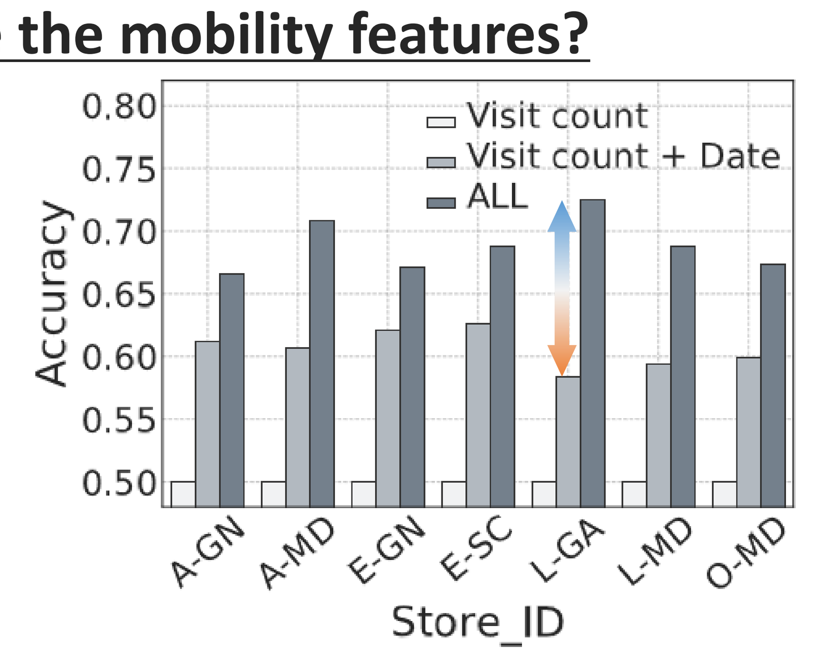
(b) All visitors: Indifferent to events.

Using our feature set, we can even predict the revisits of first-time customers with 64-72% accuracy. We test our model on seven stores. Below figures illustrate our main findings. ☺

1. How effective are the mobility features?

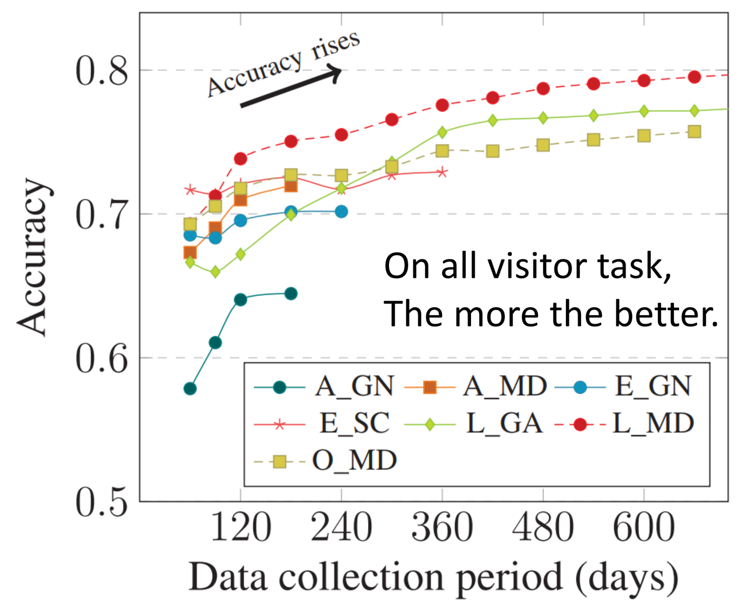


(a) On all visitors.

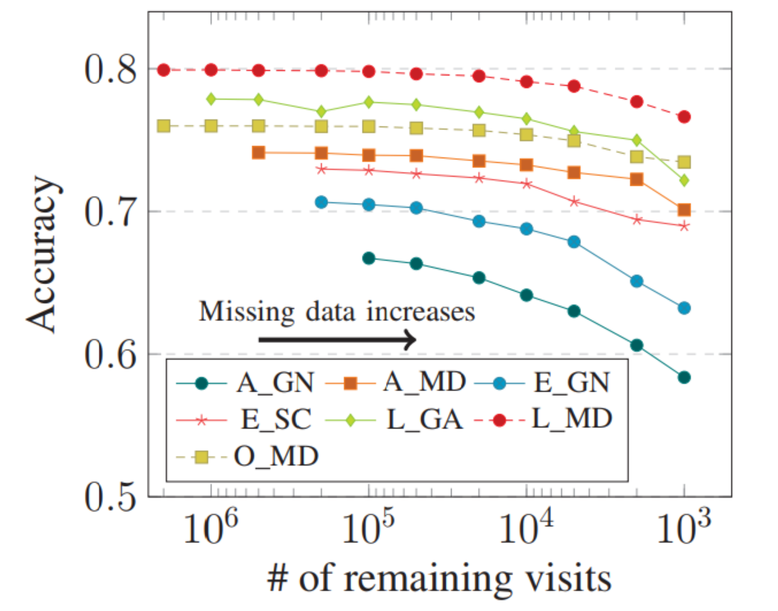


(b) On first-time visitors.

2. How much data do we need to study revisit?



3. Is our model robust to missing customers?



Statistics of our datasets.

Shop ID	A_GN	A_MD	E_GN	E_SC	L_GA	L_MD	O_MD
Length (days)	222	220	300	373	990	747	698
# sensors	16	27	40	22	14	11	27
Data size	15GB	77GB	148GB	99GB	164GB	242GB	567GB
# visits > 60s	0.11M	0.33M	0.18M	0.27M	1.06M	1.72M	2.01M
Avg. revisit rate	11.73%	31.99%	21.18%	36.55%	21.22%	32.98%	48.73%

Description of the representative features. →

Data sources	Feature groups	Twenty representative features (Among 866 features of store E_GN)	Semantic level of features					
			Sensor	Category	Floor	Gender	In/Out	None
Moving pattern of the visit	Overall statistics **[OS]** (IV-A1)	f1 = Total dwell time					✓	
		f2 = Trajectory length	✓	✓	✓	✓		
		f3 = Skewness of dwell time of each area	✓	✓		✓		
	Travel Distance/ Speed/Acceleration **[TS]** (IV-A2)	f4 = Total distance traveled inside the store		✓				
		f5 = Speed based on transition time	✓	✓	✓	✓		
		f6 = First-k HWT coefficients of acceleration	✓	✓	✓	✓		
	Area preference **[AP]** (IV-A3)	f7 = Coherency of dwell time for each level		✓	✓	✓		
		f8 = Top-k-area dwell time	✓	✓	✓	✓		
	Entrance and Exit pattern **[EE]**	f9 = Exit gate	✓					
		f10 = Number of previous re-entry on that day					✓	
	Heuristics **[HR]**	f11 = Wears clothes but does not buy		✓				
	Statistics of each area **[ST]**	f12 = Number of time sensed in the area	✓	✓	✓	✓		
		f13 = Stdev of dwell time for the area	✓	✓	✓	✓		
Temporal information of the visit	Time of visit **[TV]**	f14 = Day of the week						✓
	Upcoming events **[UE]** (IV-A8)	f15 = Remaining day until the next sale						✓
		f16 = Number of holidays for next 30 days						✓
Occurrences before the visit	Store accessibility **[SA]** (IV-A9)	f17 = Number of days since the last access					✓	
		f18 = Average interarrival time					✓	
Simultaneous visits	Group movement **[GM]** (IV-A10)	f19 = Presence of companions					✓	
		f20 = Number of companions					✓	