Ph.D. Thesis Defense

Revisit Prediction Using Customer Mobility Data

Sundong Kim

Graduate School of Knowledge Service Engineering

KAIST

May 16, 2019

Thesis Committee

- Prof. Jae-Gil Lee (Chair)
- Prof. Mun Yong Yi
- Prof. Kyoung-Kuk Kim
- Prof. Young-Jae Jang
- Prof. Meeyoung Cha



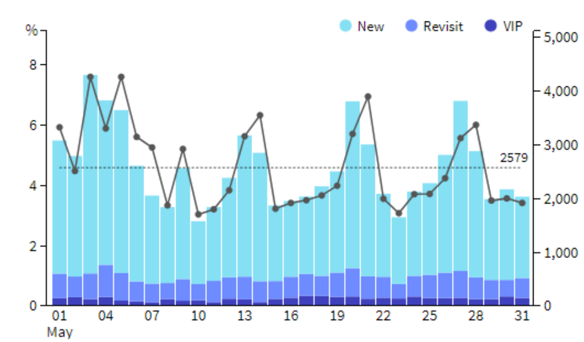

Capturing Customer Mobility

Image courtesy of Walkinsights.

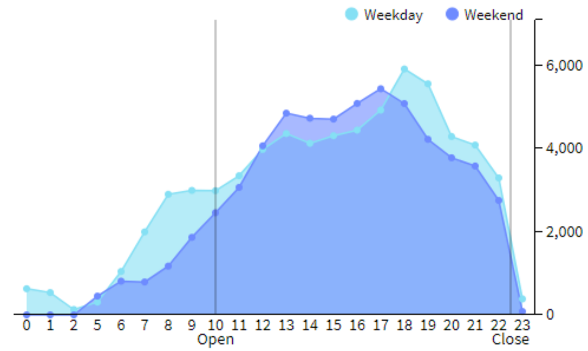
What Retail Analytics Do & Want

- Provide a dashboard, as well as consultancy services
- Use the collected information to change the store

Visitors by date



Outside traffic by hour



Ours vs. other stores

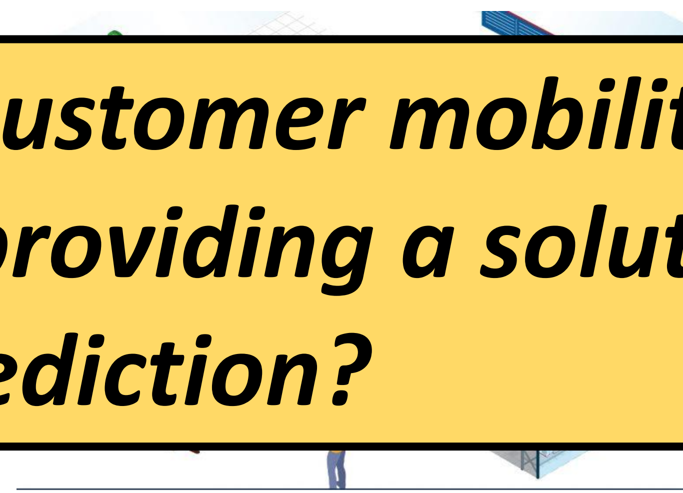
- To increase the long-term profit and revenue
- To increase the customer lifetime value

→ Securing new customers + **Keep existing customers**

Revisit Prediction

- **Retaining customers** is very important. (5% → 25-95%($), 65% rule)

- However, more than **70%** of visits are from first-time visitors and their revisit rate is only **15%.**



We have monitored customer mobility. Can we contribute by providing a solution to revisit prediction?

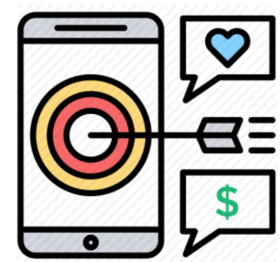
Application in Business

"Revisit Prediction for **Targeted Marketing**"

Potential to revisit?

High → Invite to premium brand (Upselling)

Low → Discount or Mass sale



- Expect to observe higher customer lifetime value.
- Feasible strategy if a company has a whole pipeline from data acquisition to marketing service

→ Knowing the potential characteristics is very important.

Thesis Goal and Focus

To Discover the Relation between
Customer Revisit and their Mobility

Our Focus: To Better Predict **Customer Revisit** by

- T1. Finding Effective Attributes
  by **Feature Engineering**
- T2. Handling Partial Observations
  by **Deep Survival Analysis**

Revision Summary

- Provide preprocessing details (pp. 12–15)
- Strengthen related work: (pp. 9–12, pp. 49–53)
- Comparison between diverse exp. settings: (p. 29)
- Model parameters: (p. 35)
- Exp. on data collection period: (pp. 39–40)
- Provide reasoning on first-time visitors: (p. 42)
- **New methodology: (pp. 46–67)**
- Exploratory data analysis: (pp. 71–76)

Outline

- Introduction
- **T1. Revisit Prediction By Feature Engineering <<**
- T2. Revisit Prediction By Deep Survival Analysis
- Conclusion

ICDM'18, KAIS J

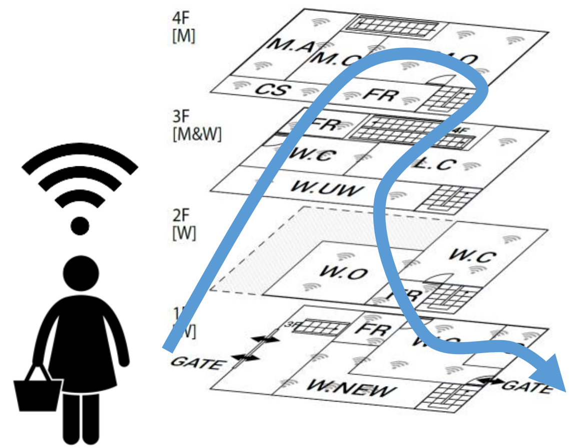
T1. Feature Engineering

" To find the **effective attributes** to determine **customer revisit** from their **mobility** "

- **Data:** Customer mobility data captured in seven stores
- **Findings:**
  - Ten groups of handcrafted features ✓
  - Performance improvement by utilizing indoor trajectories ✓
  - Predictive powers of each feature groups
  - Predictive powers by collecting longer period
  - Robustness on missing data ✓
  - LGB—Fast and high performance classifier ✓

Mobility Data from In-Store Sensors

- 7 Flagship stores
- 110K-2M visits/store
- 220-990 days collected
- Avg. traj length = 6.56



Shop ID	A_GN	A_MD	E_GN	E_SC	L_GA	L_MD	O_MD
Location	Seoul, Korea						
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
Revisit rate	11.73%	31.99%	21.18%	36.55%	21.22%	32.98%	48.73%

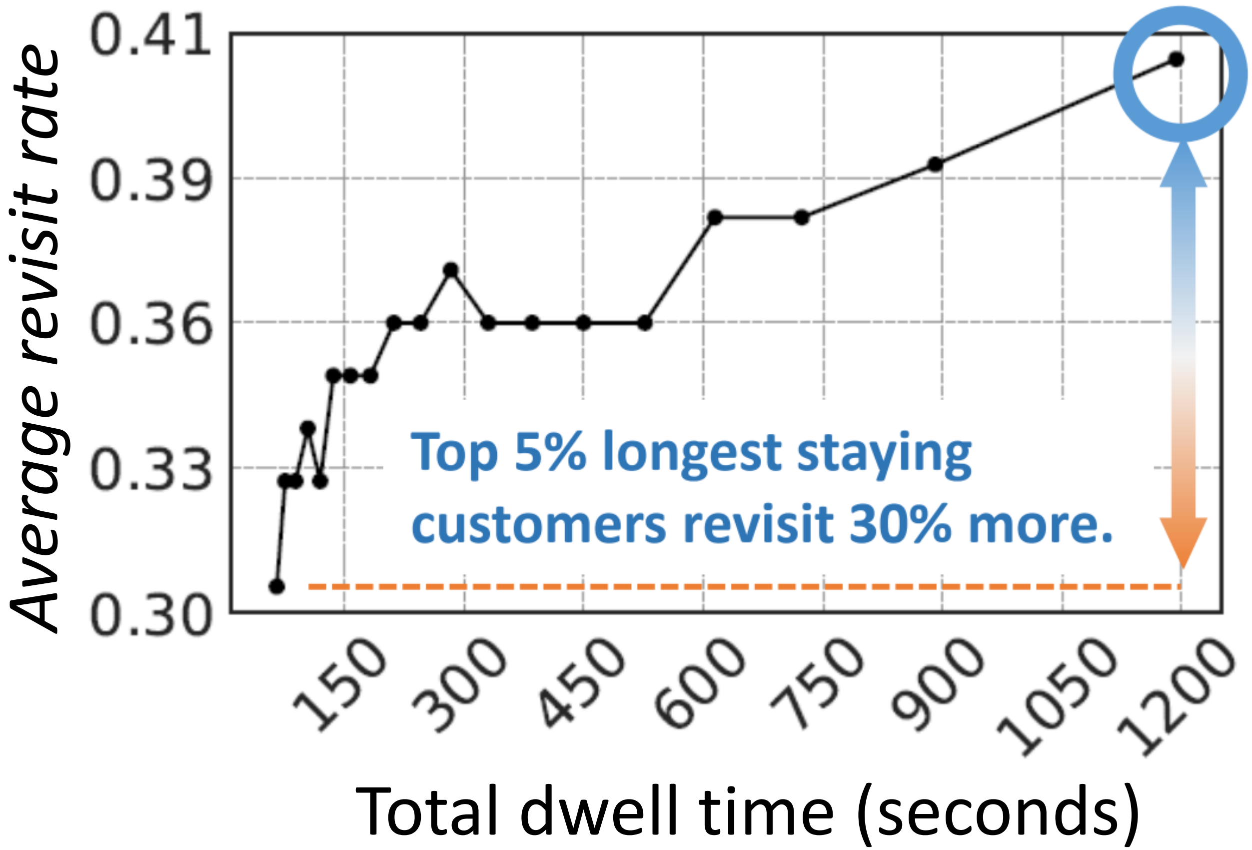
Feature Groups

- Overall statistics ✓
- Travel distance/speed/acceleration
- Area preference
- Entrance and exit pattern
- Heuristics
- Statistics of each area
- Store accessibility ✓
- Group movement
- Time of visit
- Upcoming events ✓

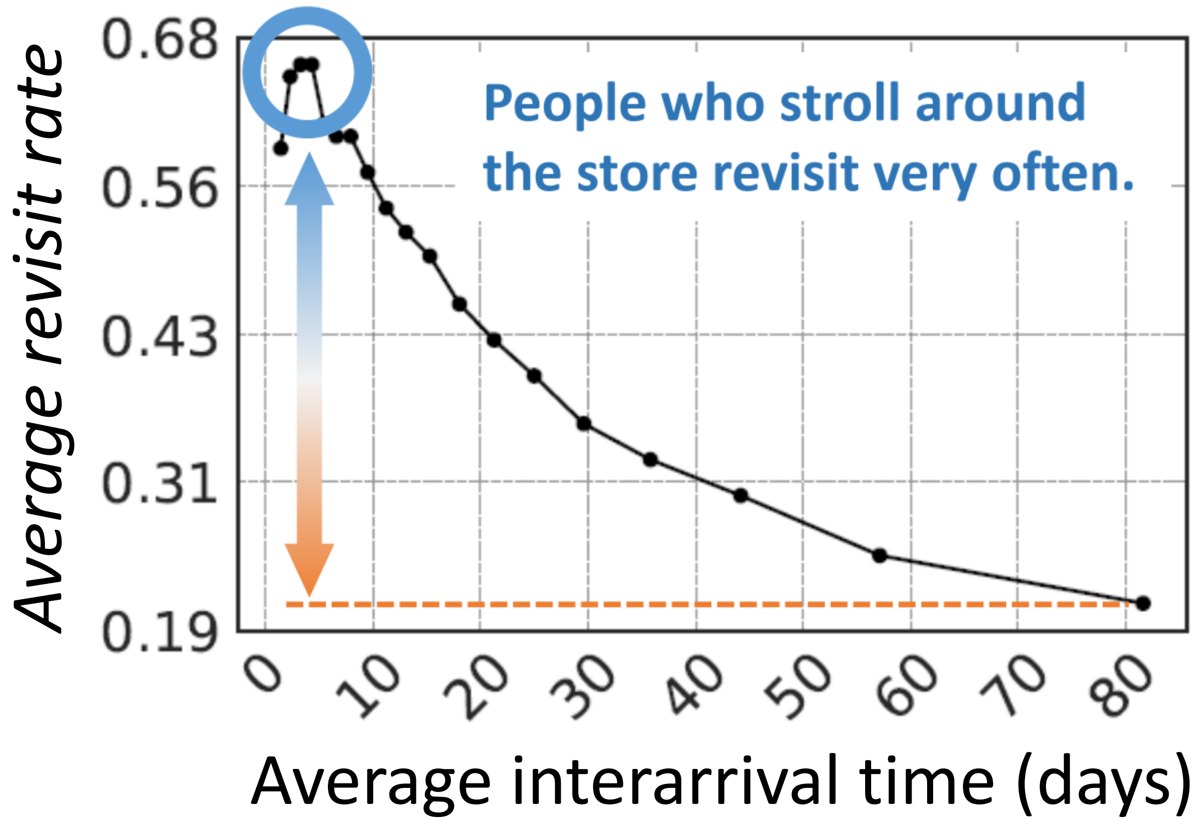
Motion pattern

Temporal Information

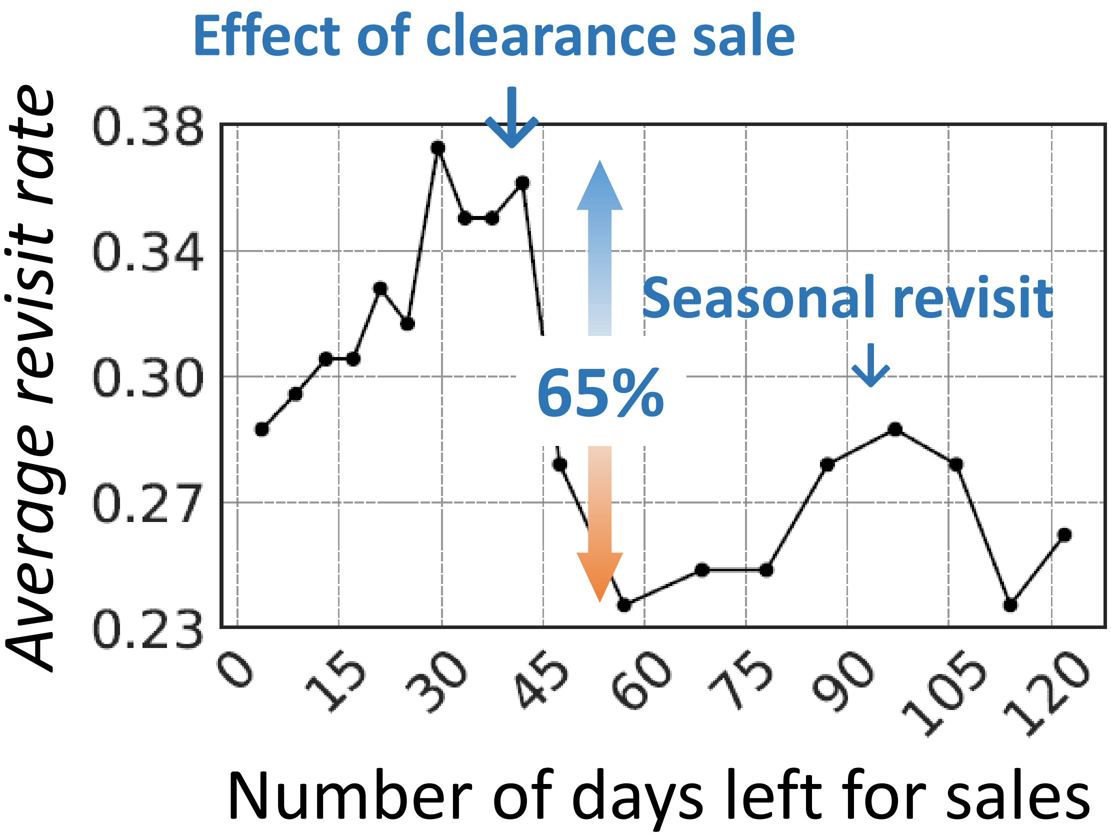
Total Dwell Time



Store Accessibility



“Sale” for First-Time Visitors

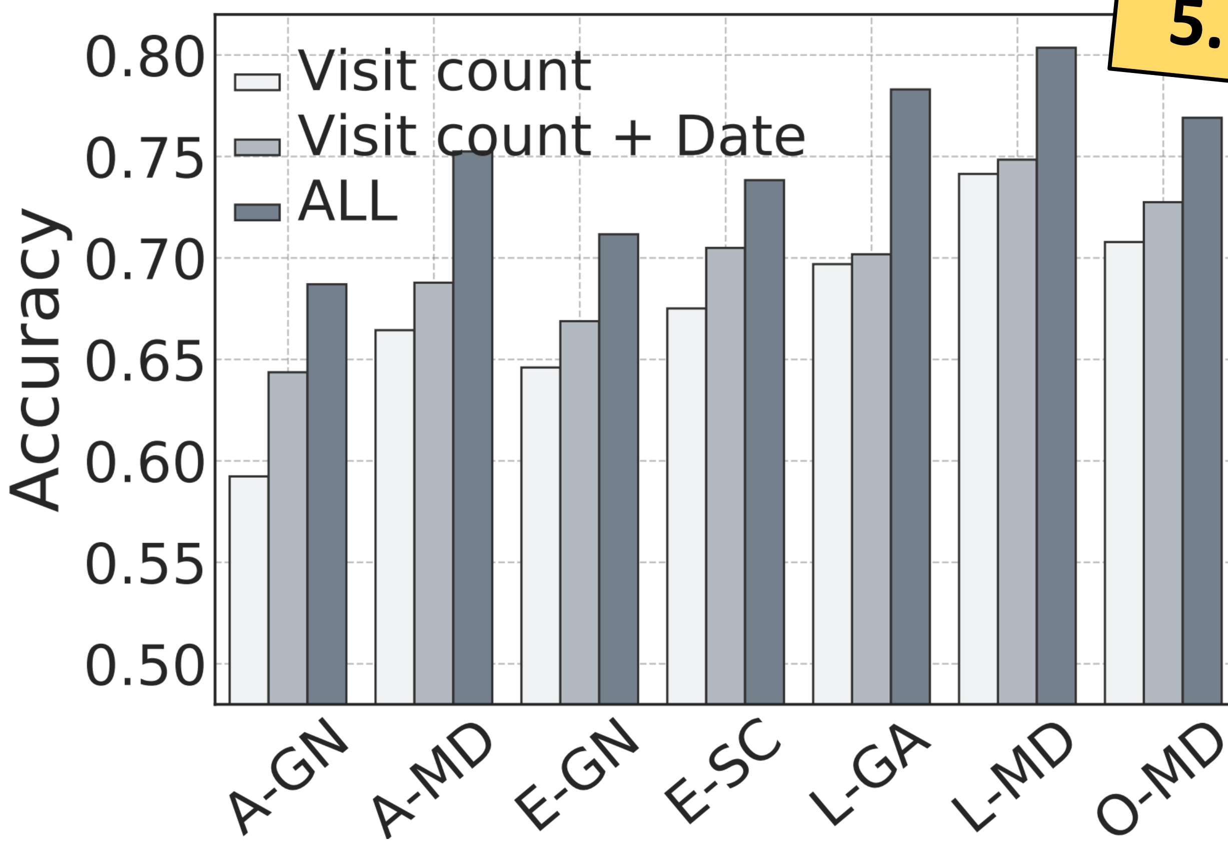


On Customer Groups

Store ID	A_GN	A_MD	E_GN	E_SC	L_GA	L_MD	O_MD
# visits							
v_1	0.661	0.741	0.681	0.716	0.763	0.778	0.758
v_2	0.732	0.735	0.716	0.691	0.795	0.773	0.706
v_3	0.824	0.786	0.791	0.751	0.840	0.848	0.757
v_4	0.856	0.808	0.845	0.800	0.848	0.879	0.801
v_5	-	0.803	0.865	0.831	0.847	0.885	0.820
v_6	-	0.810	0.884	0.852	0.846	0.883	0.829
v_7	-	0.808	0.907	0.861	0.856	0.879	0.834
v_8	-	0.814	0.911	0.866	0.836	0.878	0.838
v_9	-	0.802	0.903	0.875	0.863	0.874	0.837
v_{10}	-	0.789	-	0.900	0.867	0.870	0.839

New contents

Comparison with Baselines



Improved by 5.1-12.4%.

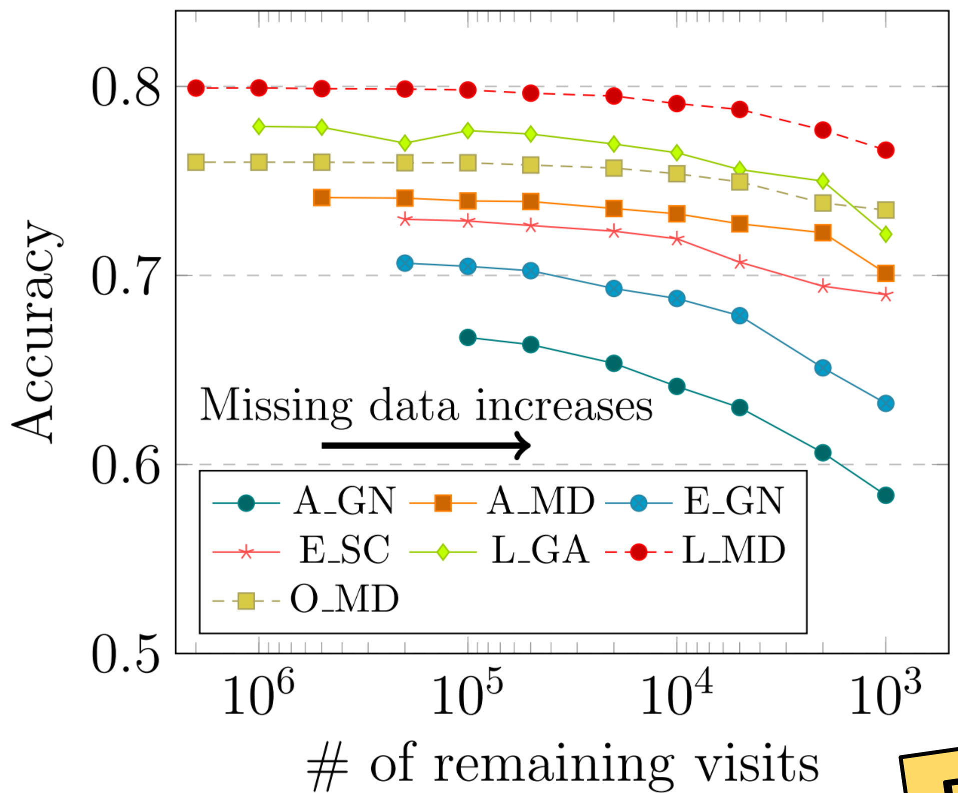
Comparison with Baselines on v_i

Store ID	A_GN	A_MD	E_GN	E_SC	L_GA	L_MD	O_MD
# visits							
v_1	18.6/7.7	17.1/14.7	12.9/9.1	10.4/7.1	18.2/17.6	10.5/10.4	7.6/7.4
v_2	4.9/1.2	13.5/5.0	7.5/2.0	15.1/3.1	4.6/3.0	18.4/12.5	29.7/13.0
v_3	1.7/0.4	4.2/1.3	3.0/0.4	7.5/1.3	0.9/0.3	2.5/1.2	8.0/3.5
v_4	1.3/0.3	3.5/0.5	2.8/1.1	5.5/0.7	1.0/0.1	0.9/0.2	3.7/1.0
v_5	-	3.2/0.3	1.3/-0.4	3.8/0.8	1.1/0.1	0.7/0.0	2.7/0.5
v_6	-	2.3/0.2	1.6/0.8	3.3/0.4	1.3/0.2	0.8/0.0	2.4/0.2
v_7	-	3.8/0.8	1.8/-0.1	2.7/1.0	1.3/0.3	0.8/0.0	2.2/0.2
v_8	-	4.0/-0.2	1.7/0.5	2.4/0.0	1.4/0.2	1.2/0.0	2.2/0.2
v_9	-	3.6/0.0	1.5/0.9	3.2/0.6	1.8/0.6	1.4/0.2	2.0/0.0
v_{10}	-	3.1/0.0	-	2.1/0.2	0.9/0.2	1.6/-0.1	2.5/0.2

Revised contents

Robustness on Missing Customers

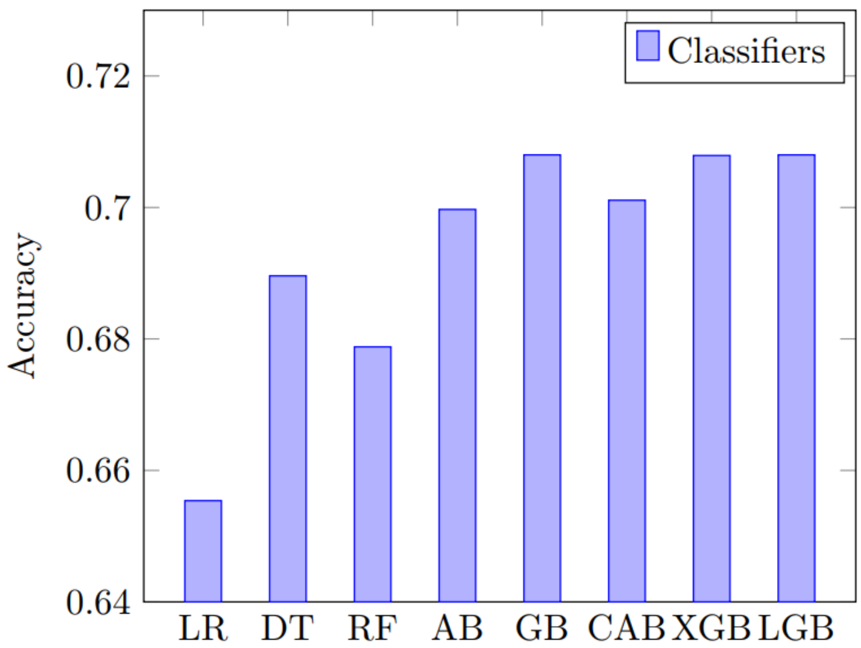
- Over 95% of the performance is maintained with a very small fraction of the dataset (e.g., **0.5%** for L_MD)



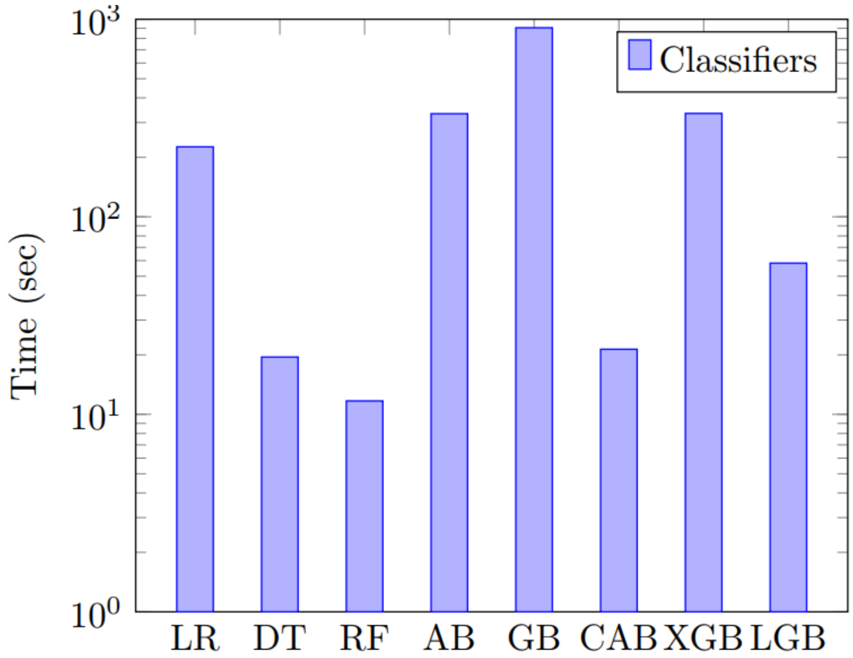
Revised contents

Comparison Between Classifiers

1) Prediction accuracy



2) Execution time



New contents

Stacking Options

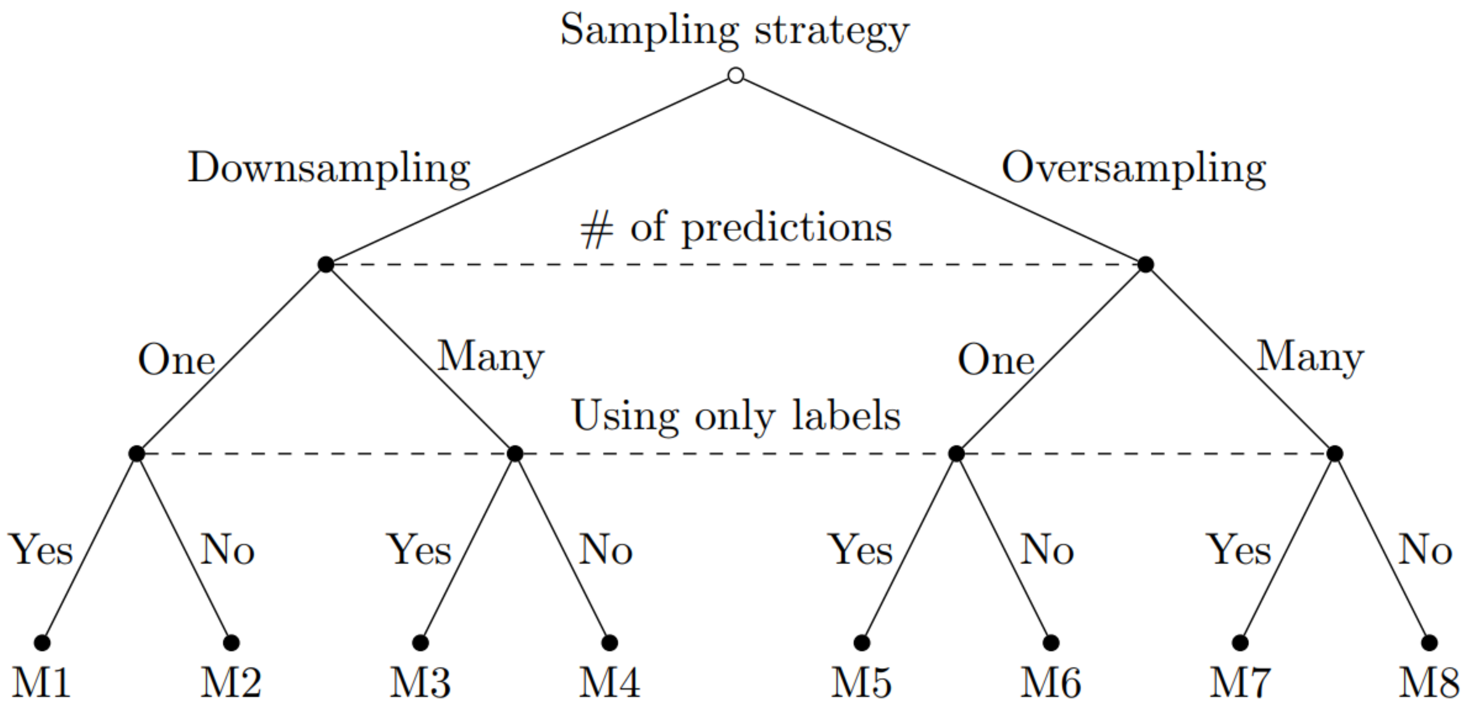


Figure 3.7: Stacking options.

New contents

Outline

- Introduction
- T1. Revisit Prediction By Feature Engineering
- **T2. Revisit Prediction By Deep Survival Analysis <<**
- Conclusion

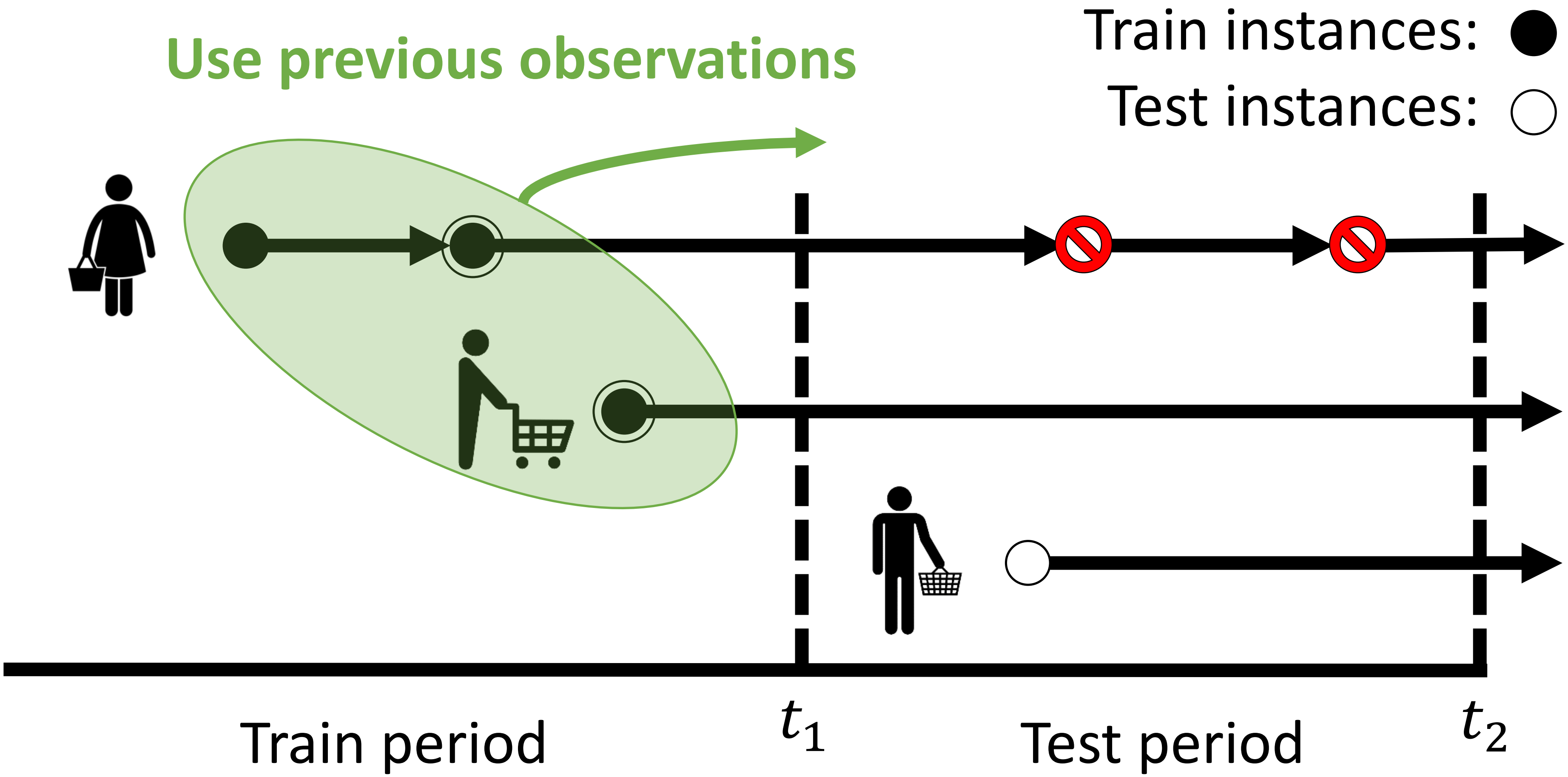
Work after proposal

T2. Deep Survival Analysis

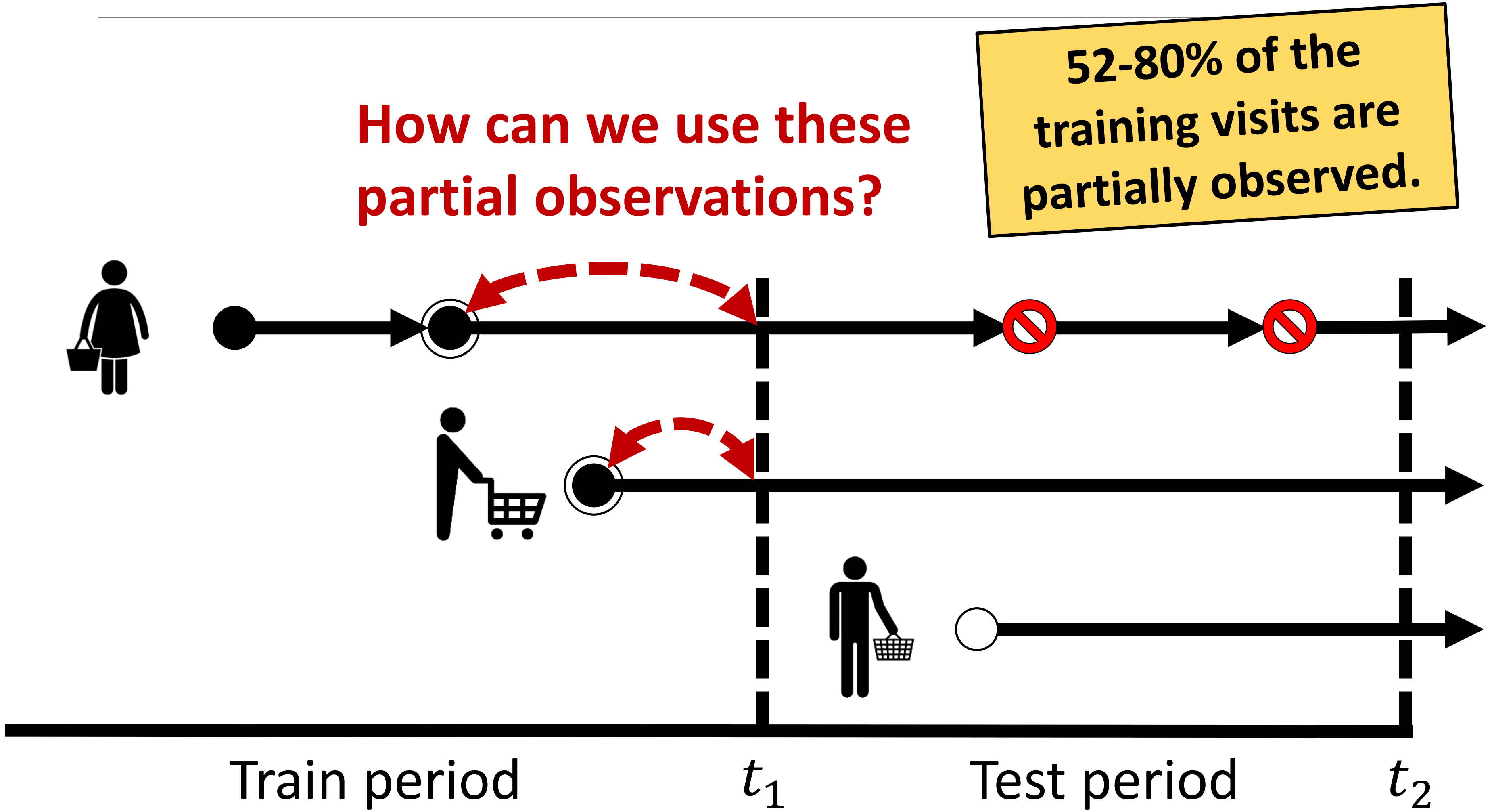
" To make the most of every observation
by **deep survival analysis**"

- **Focus:** Handling **partial observations**
  - Have not been used for the regression task
  - Cause imbalanced class distribution
  - Include all first-time visitors, large portion of the datasets
- **Findings:**
  - Develop a model ***SurvRev*** powered by **Survival analysis** and **Deep learning**
  - Implement two modules to encode the visit and histories
  - Optimize custom loss functions to meet multiple objectives

Prediction on Longitudinal Data



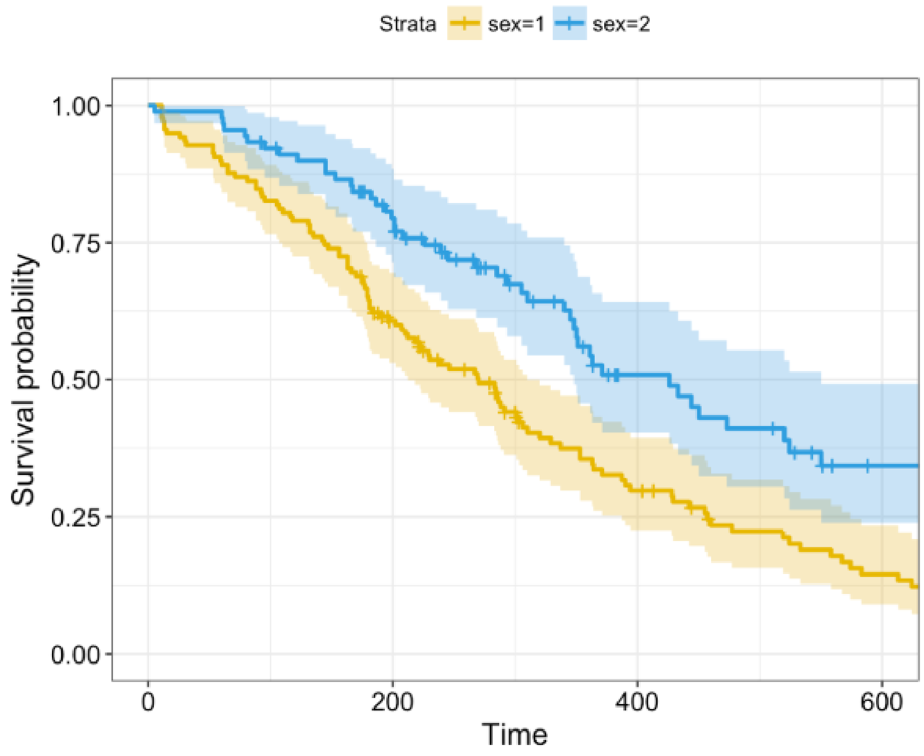
Motivation



52-80% of the training visits are partially observed.
How can we use these partial observations?
Train period
t_1
Test period
t_2

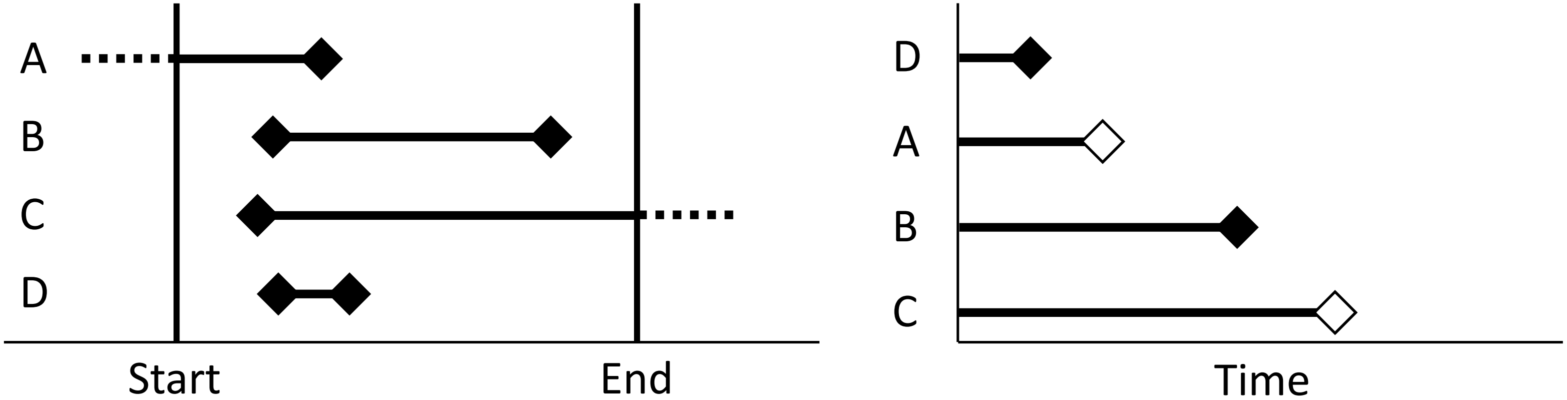
Survival Analysis – Purpose

- To analyze the expected duration of time until one or more event happen, such as death, failure, marriage, next visit, etc.



Survival Analysis

- **Censoring:**



- **Analysis Issue:**
  - If there is no censoring, standard regression procedures could be used
  - However, time to event is restricted to be positive, skewed distribution

Survival Analysis – Notation

- **Survival function**

$$S(t) = \Pr(T > t) = 1 - F(t)$$

- **Hazard function (= Event rate, Revisit rate)**

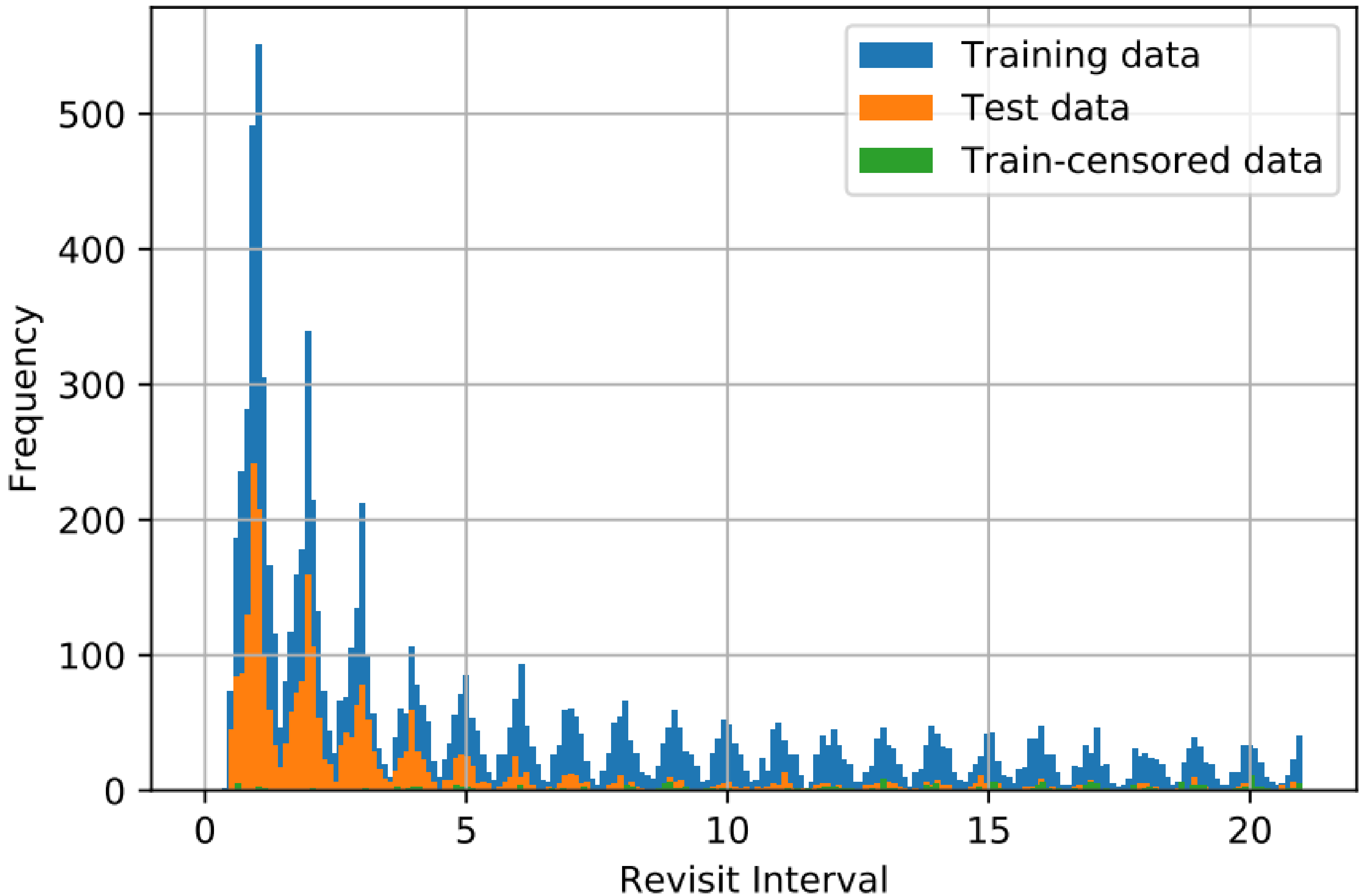
$$\lambda(t) = \lim_{dt \to 0} \frac{\Pr(t \le T < t + dt \mid T > t)}{dt} = \frac{f(t)}{S(t)}$$

$f(t)$: Event density function
$F(t)$: Cumulative distribution function

Estimating S(t) and $\lambda(t)$

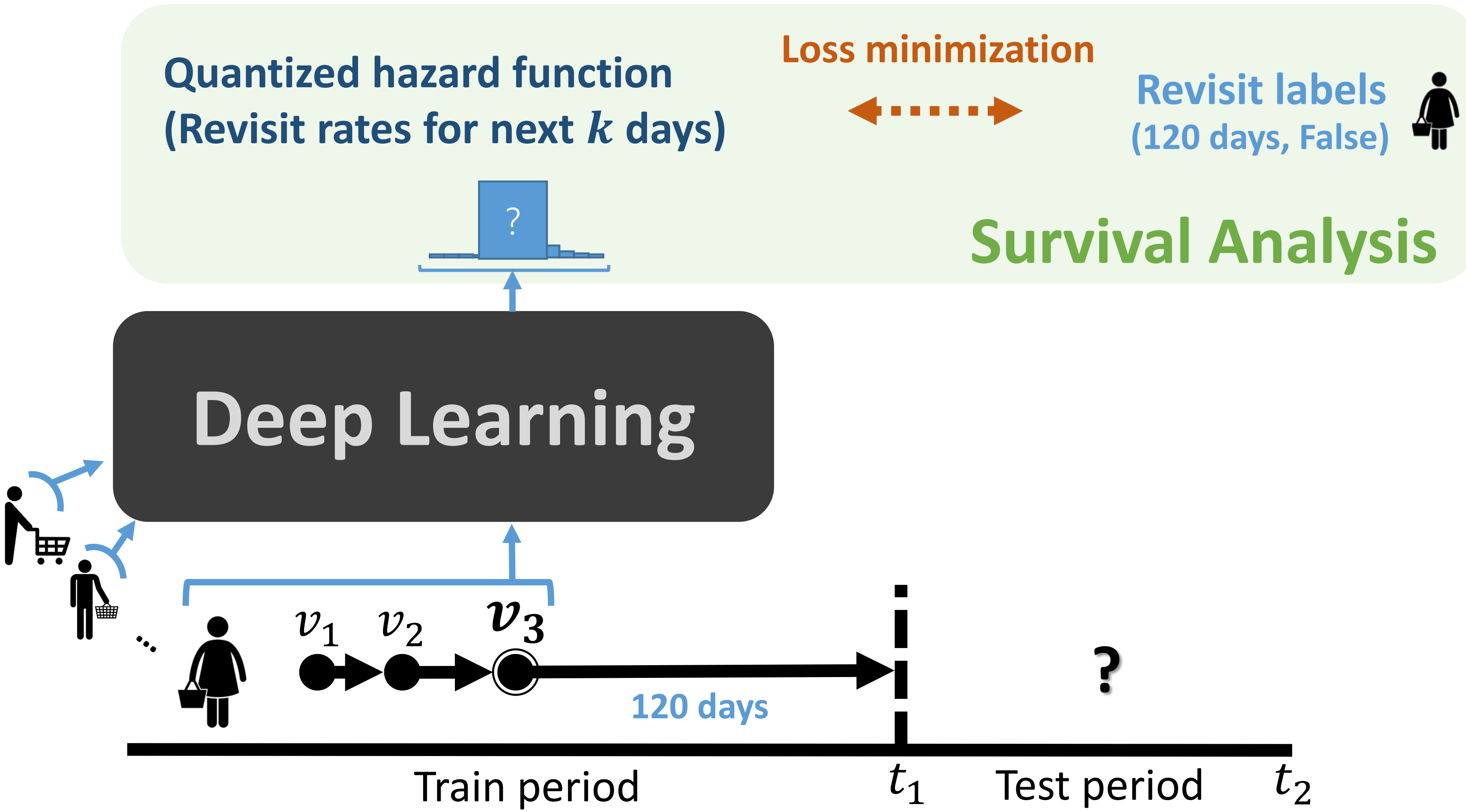
- Nonparametric estimators [Kaplan1958]:
  - If every subject follows the same survival function
- Parametric estimators:
  - Exponential
  - Weibull
- Semi-parametric estimators [Cox1972]:
  - $\lambda(t|\boldsymbol{x}) = \lambda_0(t)e^{\beta\boldsymbol{x}}$
  - The base hazard function has some assumption, e.g., Weibull distribution.

→ **Drawback: Not flexible in practice.**

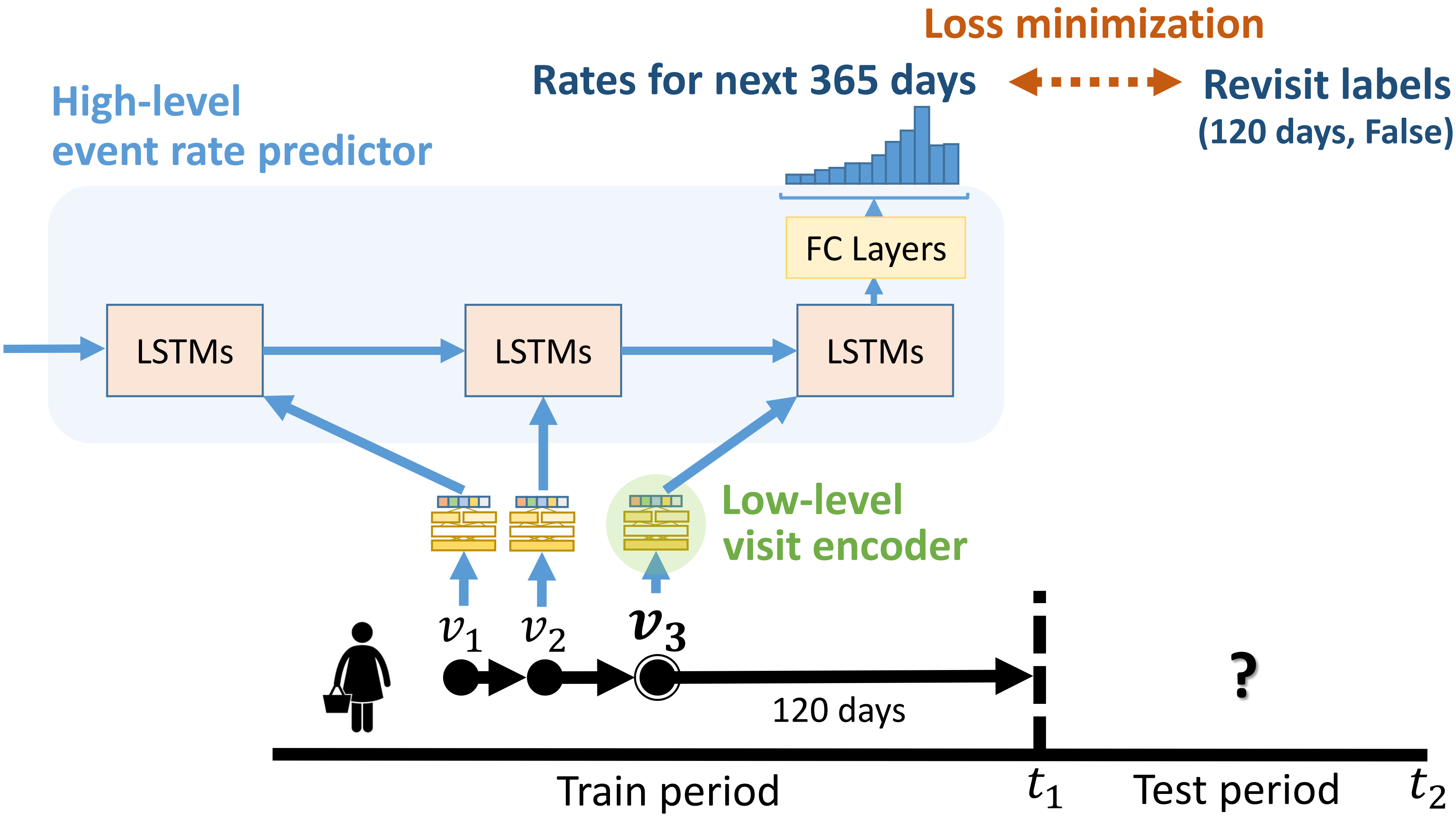


→ Drawback: Not flexible in practice.

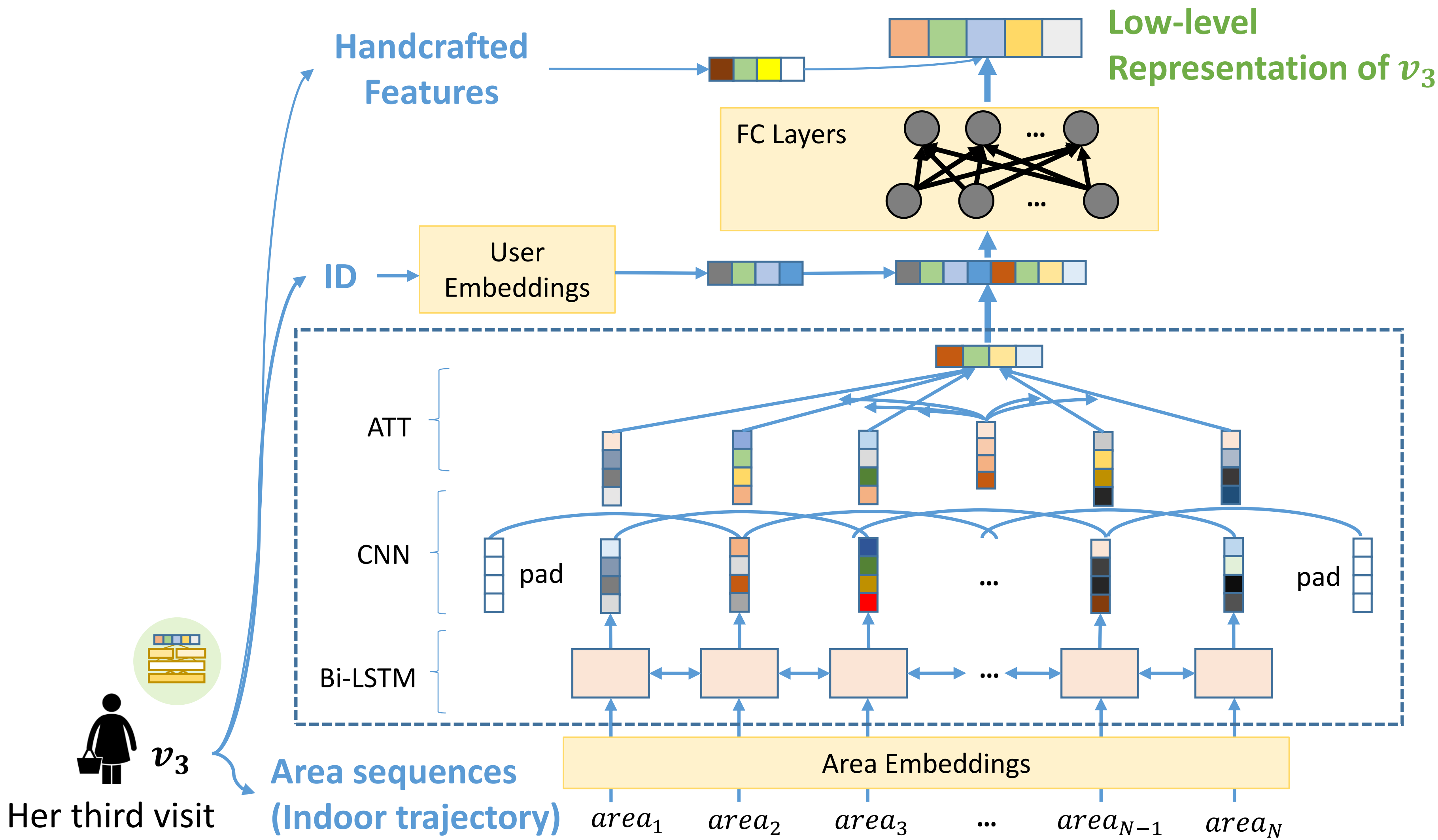
How to Use SA and DL?



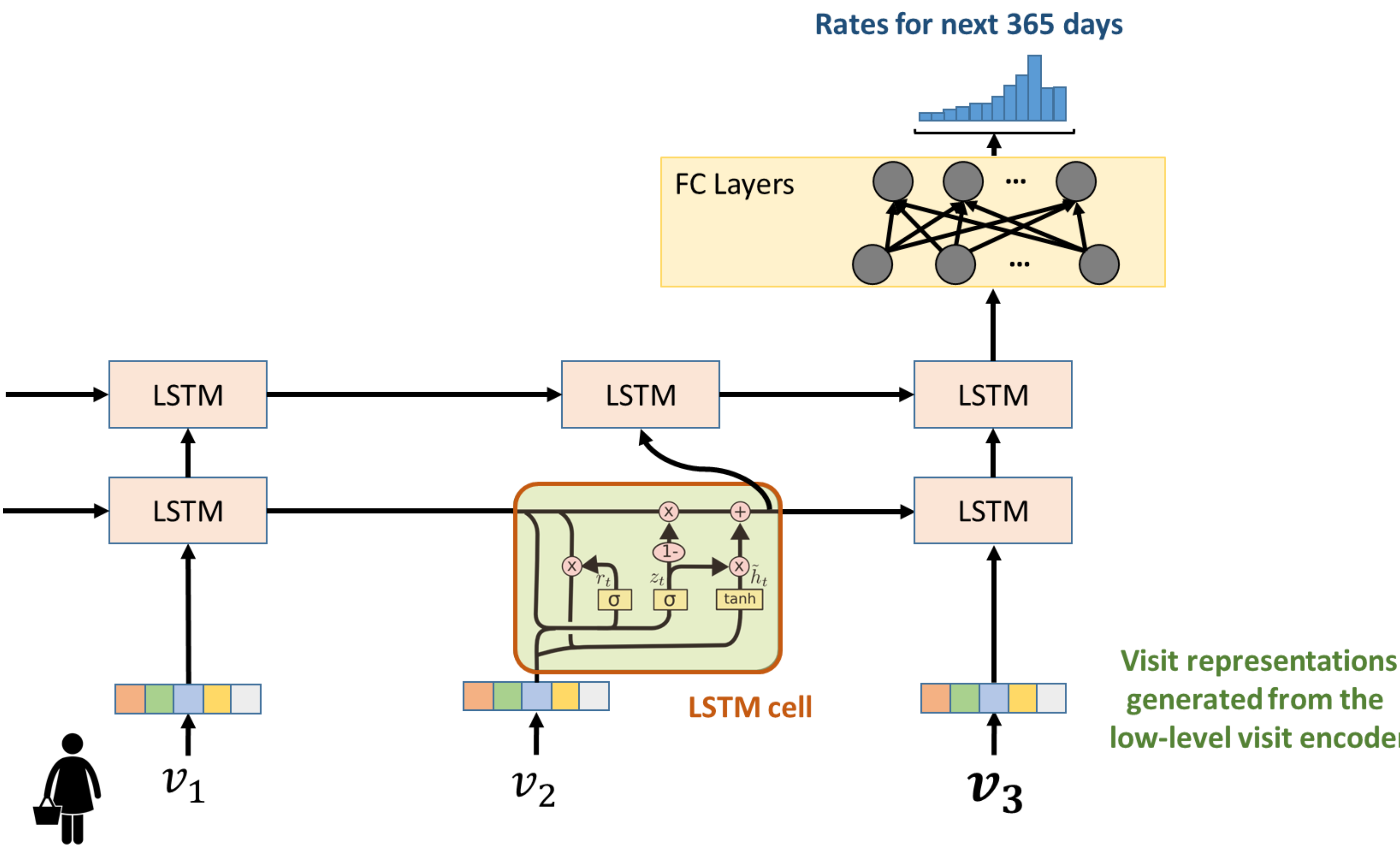
SurvRev Architecture



Low-level Encoder



High-Level Event Rate Predictor



Loss Function

$$\boldsymbol{\mathcal{L} = \mathcal{L}_{nll} \cdot \mathcal{L}_{rmse} \cdot \mathcal{L}_{ce} \cdot \mathcal{L}_{rank}}$$

- $\mathcal{L}_{nll}$: Negative log-likelihood loss [Ren2019]
- $\mathcal{L}_{rmse}$: RMSE loss [Kim2018]

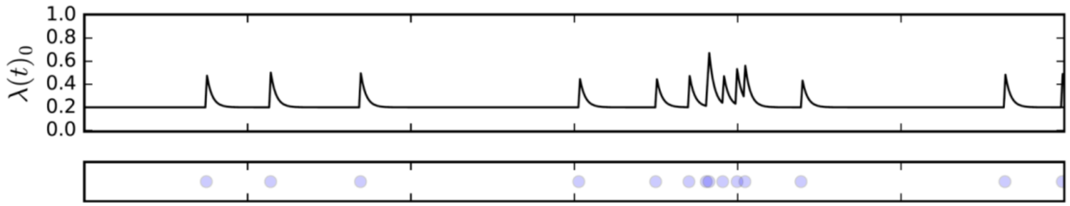
} For uncensored cases

- $\mathcal{L}_{ce}$: Cross-entropy loss [Ren2019]
- $\mathcal{L}_{rank}$: Pairwise ranking loss [Lee2018]

} For all cases

Related Work & Baselines (1)

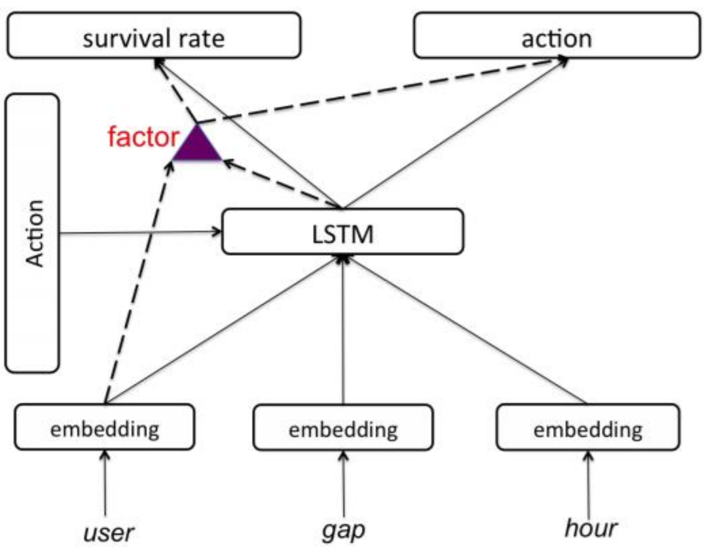
- **Majority:**
  - Prediction results follow the majority label or average value
- **Poisson Process:**
  - Interarrival times follows the exponential distribution
- **Hawkes Process [Hawkes1971]:**
  - Self-stimulating, exponentially decaying point process

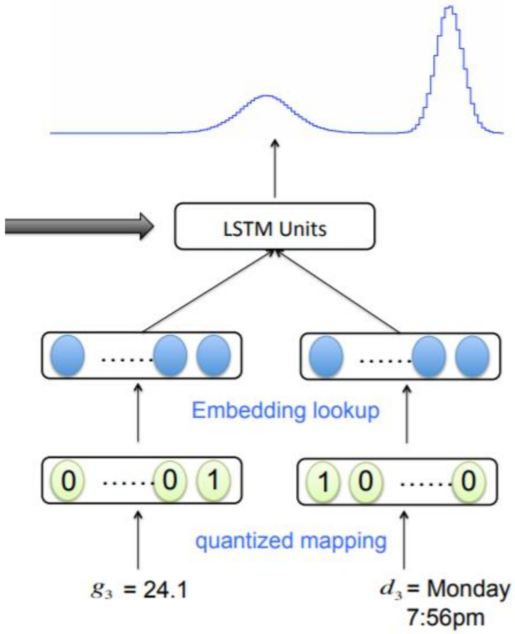


- **Cox Proportional Hazard Model [Cox1972]**
  - Semi-parametric statistical model

Related Work & Baselines (2)

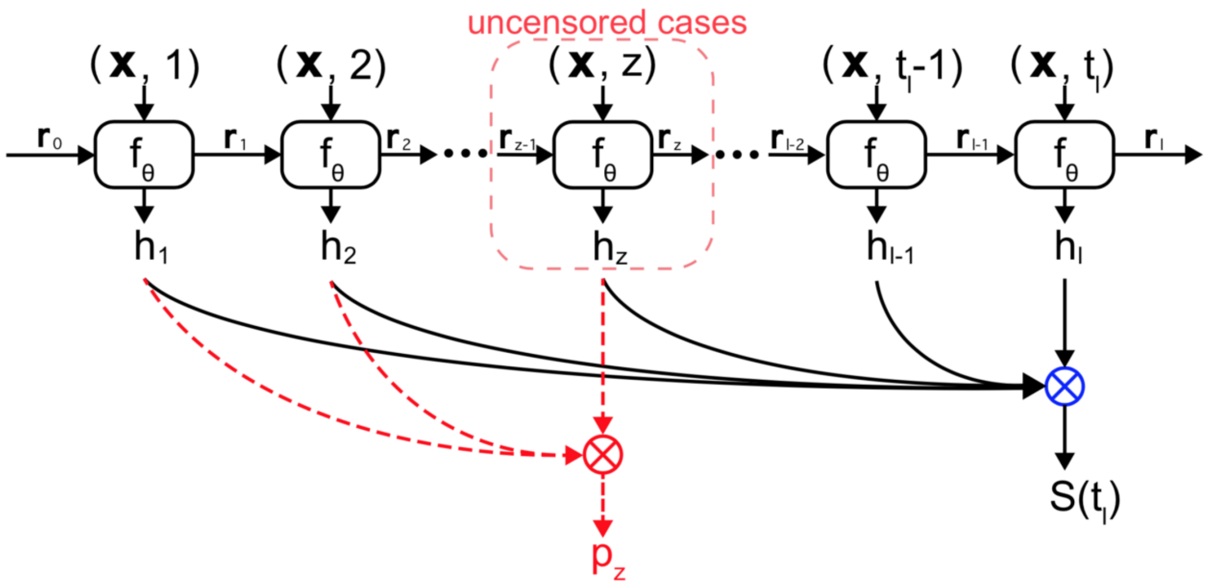
- **Neural Survival Recommender [Jing2017]:**
  - **A deep multi-task learning model with LSTM and 3-way factor unit, used for churn analysis in music streaming**
  - **Did not consider lower-level interactions**





Related Work & Baselines (3)

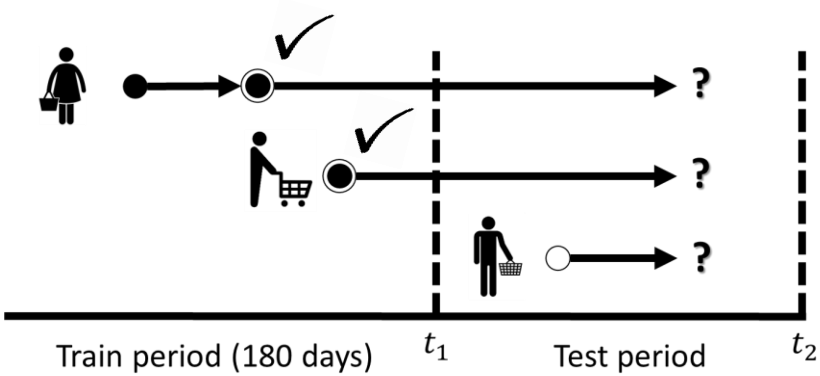
- **Deep Recurrent Survival Analysis [Ren2019]**
  - **An auto-regressive model**
  - **Each cell emits a hazard rate**
  - **Each LSTM considers only a single event**



Evaluation on Censored Customers

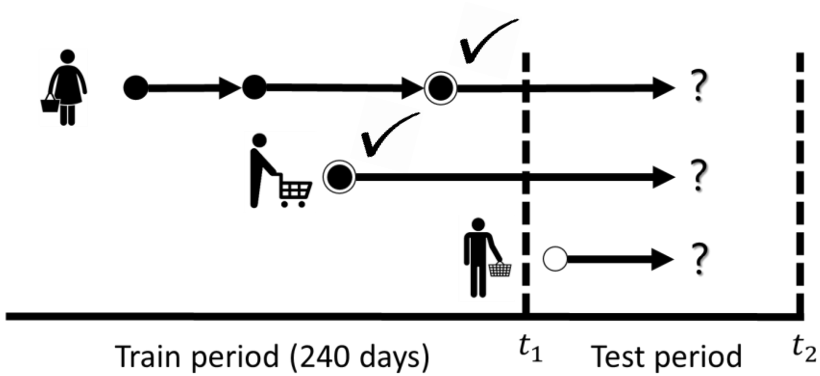
C-index results (180 days).

	Store A	Store B	Store C	Store E
Majority	0.500	0.500	0.500	0.500
Poisson	0.528	0.591	0.588	0.582
Hawkes	0.530	0.593	0.588	0.580
XGB	0.420	0.597	**0.671**	0.549
NSR	0.497	0.497	0.480	0.523
DRSA	0.500	0.500	0.499	0.500
SurvRev	**0.561**	**0.672**	0.649	**0.647**



C-index results (240 days).

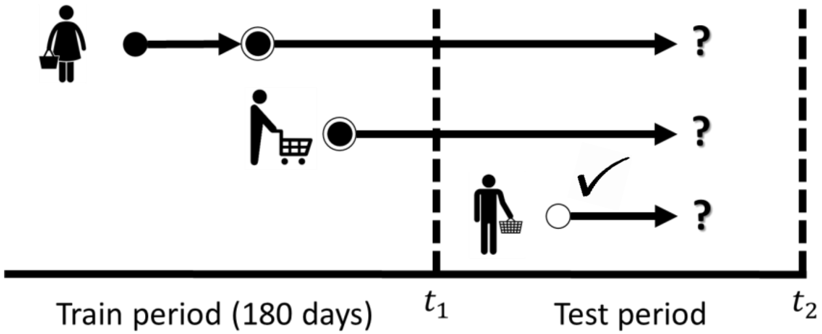
	Store A	Store B	Store E
Majority	0.500	0.500	0.500
Poisson	0.552	0.622	0.617
Hawkes	0.549	0.624	0.613
XGB	**0.667**	0.568	**0.830**
NSR	0.509	0.513	0.504
DRSA	0.500	0.500	0.501
SurvRev	0.606	**0.726**	0.702



Evaluation on New Customers

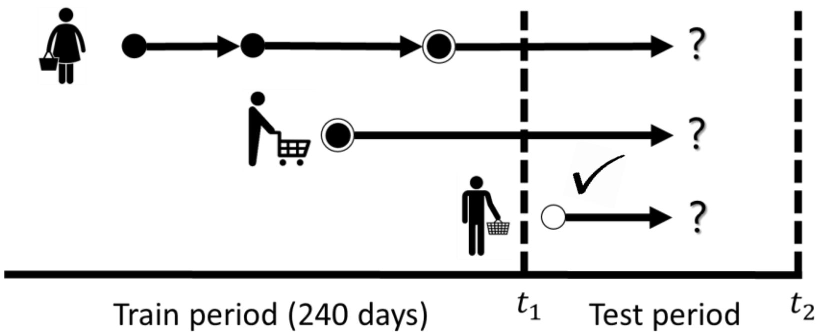
F-score results (180 days).

	Store A	Store B	Store C	Store E
Majority	0.000	0.000	0.000	0.000
Poisson	0.244	0.302	0.415	0.244
Hawkes	0.242	0.304	0.412	0.241
Cox-ph	0.286	0.353	0.176	0.000
XGB	0.236	0.317	0.248	0.097
NSR	0.000	0.000	0.000	0.000
DRSA	0.298	0.360	**0.461**	0.277
SurvRev	**0.315**	**0.373**	0.458	**0.295**



F-score results (240 days).

	Store A	Store B	Store E
Majority	0.000	0.000	0.000
Poisson	0.214	0.275	0.204
Hawkes	0.212	0.276	0.209
Cox-ph	0.000	0.000	0.000
XGB	0.025	0.194	0.000
NSR	0.000	0.000	0.000
DRSA	0.245	0.300	0.223
SurvRev	**0.272**	**0.307**	**0.263**

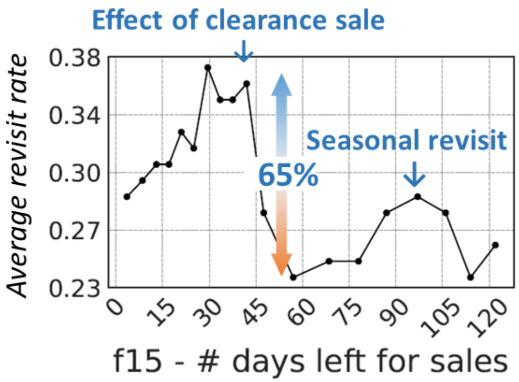


Outline

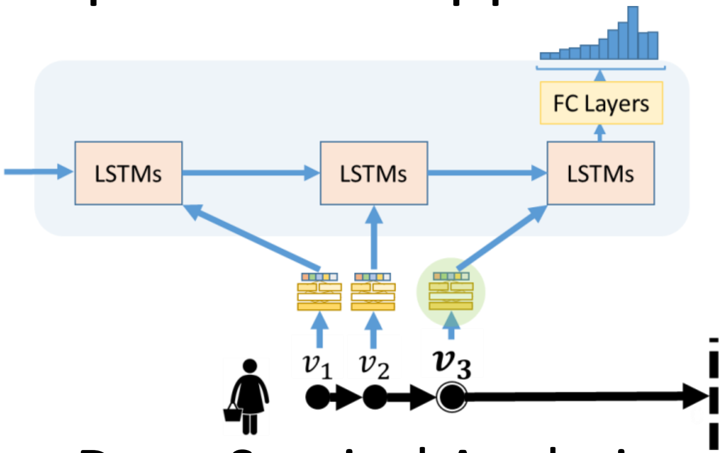
- Introduction
- T1. Revisit Prediction By Designing Features
- T2. Revisit Prediction By Designing a Model
- **Conclusion <<**

Conclusion

- **Goal:** To discover the relation between **Customer Revisit** and their **Mobility**
- **Contributions:** Developed two prediction approaches



Feature Engineering



Deep Survival Analysis

- **Impact:**

Reference

- **[Kim2018]** S. Kim and J.-G. Lee, “Utilizing in-store sensors for revisit prediction," in *IEEE International Conference on Data Mining*. IEEE, 2018, pp. 217–226.
- **[Kim2019]** S. Kim and J.-G. Lee, “A systemic framework of predicting customer revisit with in-store sensors," in *Knowledge and Information Systems (To Appear)*. Springer, 2019.
- **[Kaplan1958]** E. L. Kaplan and P. Meier, “Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, vol. 35, pp. 457–481, 1958.
- **[Hawkes1971]:** A. G. Hawkes, “Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58, no. 1, pp. 83–90, 1971.
- **[Cox1972]:**D. R. Cox, “Regression models and life-tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 34, no. 2, pp. 187–220, 1972.
- **[Jing2017]** H. Jing and A. J. Smola, “Neural survival recommender," in *The 10th ACM International Conference on Web Search and Data Mining*. ACM, 2017, pp. 495–503.
- **[Lee2018]** C. Lee, W. R. Zame, J. Yoon, and M. van der Schaar, “DeepHit: A deep learning approach to survival analysis with competing risks," in *The 32nd AAAI Conference on Artificial Intelligence*. AAAI Press, 2018.
- **[Ren2019]** K. Ren, J. Qin, L. Zheng, Z. Yang, W. Zhang, L. Qiu, and Y. Yu, “Deep recurrent survival analysis," in *The 33rd AAAI Conference on Artificial Intelligence*. AAAI Press, 2019.

Thank you!