A deep real options policy for sequential service region design and timing

[Rath, S.](), & [Chow, J. Y](). (2022).
C2SMART University Transportation Center
Department of Civil and Urban Engineering
https://arxiv.org/abs/2212.14800

Reading Seminar #2

2023/05/09(Tue)

Fuga Mayuzumi

MY interest

Network and service design with vehicles-to-users matching and scheduling in large-scale Network

Vehicle routing problem's major requirements are often in trade-off relationship

- Demand-responsive vs. Non-myopic
- Efficiency vs. Social Welfare

▸ Evaluating feasible matching patterns needs to compared

MY Motivation for reading this paper

Optimal matching in large-scale ridesharing market requires some techniques to mitigate calculation cost for evaluating feasible patterns.

- Introducing “Search Theory” for mitigating the enumeration cost for combinatorial problem is one of my directions.

Today’s paper is tackling mitigation of enumeration by sampling based on the stochastic nature of the variables

Authors’ Motivation

Target: Optimizing the MoD(Mobility on Demand) service region and its timing under uncertainty in demand/market uncertainty.

① Introducing the concept of Real Options(RO) to evaluate the multiple interacting real options such as deferral option and network redesign option.

▸ Flexibility in decision making is considered

② Propose a new variant ”deep” real options policy using an efficient recurrent neural network (RNN) based ML method(CR−RNN policy)

▸ Sampling sequences to avoid the need for enumeration for large scale implementation

Real options

- Originally a concept for evaluating investments in corporate finance.
- The value of this flexibility is known to increase as the volatility of the underlying stochastic processes increases
  = The greater the uncertainty, the more value there is to having the flexibility to make a decision.

Classic search theory

- Stigler Model(1961) The Economics of information

The purchasing entity extracts independent prices from the price distribution F under constant search cost c. As the number of searches increases, the expected minimum price becomes smaller. It is optimal to continue extraction until this positive marginal benefit is less than the search cost, which is called optimal stopping point.

Limitation of Stigler Model：

Lack of sequentiality in decision rule： -> McCall(1965)

The buyer knows the price distribution and decides the number of finite searches in advance, but this decision rule does not use the information obtained by price search, which is not a reasonable assumption from the standpoint of trying to buy at the lowest possible price.

Sequential search and Dynamic programming

- The optimal stopping rule of a sequential search can be calculated under a dynamic economy with state variables that vary according to the Ito process.

  - Geometric Brown Motion(GBM) is the special case of Ito process, and adopted as the stochastic process because it gives a simple form for the continuation value of Bellman Equation.

  - GBM validates that $\phi(x'|x)$ satisfies first-order stochastic dominance, under which critical value uniquely split “stopping” and "continuing.”

Ref: https://core.ac.uk/download/pdf/71792194.pdf(in Japanese)

Real options

$$\text{Expanded NPV} = \text{base NPV} + \textit{option premium} \qquad (1)$$

- Expanded NPV favors the value to react flexibly to future uncertainty opposed to conventional base NPV(net present value)
- Option premium is divided into deferral premium and network redesign premium (Chow and Regan, 2011a: previous work)
- For each investment candidate project h, what has been invested before and what is to be invested at subsequent time steps can be defined in terms of an investment sequence.

Real options

- The total number of possible sequences for a portfolio of H compounded projects is $H!$
- CR policy selects the sequence that offers the highest initial option value (Chow and Regan, 2011a).
- When the number of projects (H) increases, the possible number of sequences ($H!$) increases drastically.

▸ Evaluating the CR policy value (=approximated value function) for each of the enumerated sequences becomes computationally expensive

Assumptions on uncertainty to model

- Demand are uncertain and assumed to evolve over time following a stochastic process
- Zone interactions in terms of interrelated stochastic origin-destination demand across zones.
- Heterogeneous volatility in the stochastic elements.

Sequential design illustration

Sequentially evaluating the for feasible set of order of service region and its timing requires immense enumeration

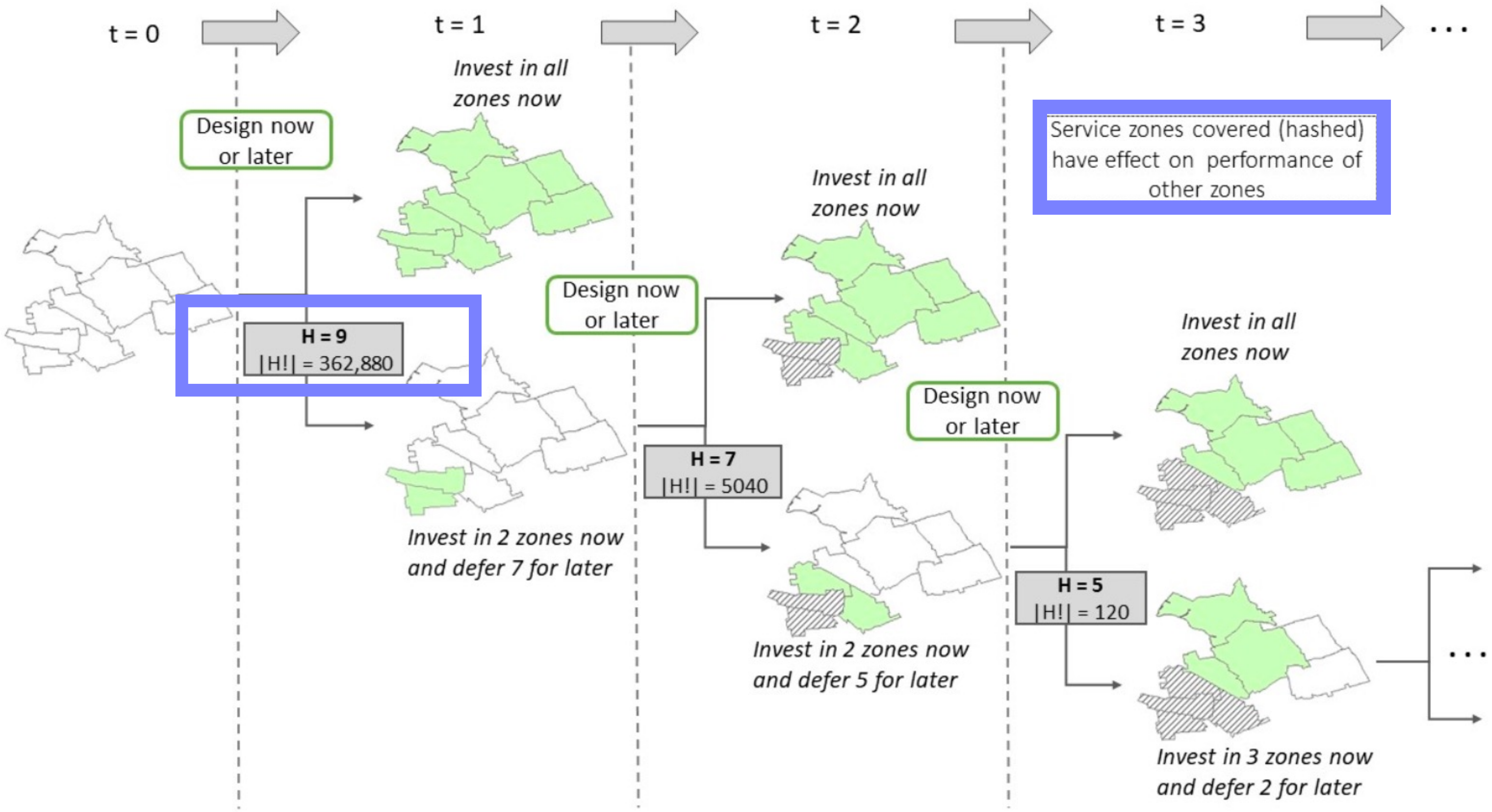


Fig. 2: Illustration of sequential service region design and timing problem (for a set of $\mathscr{H}$ zones); green highlighted service zones denote investment decisions in those zones (*i.e.*, invest immediately) at time t, hashed zones represent zones already designed (invested) in previous time steps, and no highlights at time t represent zones where it is optimal to defer the investment to a later time period.

Approach to forego enumeration

ML-based method to reinforce the method to calculate the real option value in sequential service region design and timing problem for MoD services

- Train recurrent neural network (RNN) based sequence classifier with a small (randomly sampled) fraction of sequences
- Use the trained RNN to predict top sequences (i.e., sequences with high policy values) from the large set of remaining candidate sequences.
- The RNN training involves the novel use of a gap relative to a reference policy value (Chow and Sayarshad, 2016) for avoiding full sequence evaluation

Related Literature 1/3

A. Service region design

- Cordon design for congestion measures (Zhang and Yang 2014)
- Continuous approximation model of performance within region (e.g., Daganzo and Newell 1986)
- Aggregate network models for taxi (Salanova et al. 2011,2014)

- Activity-based connected subgraph problem (Chow 2021)

Related Literature 2/3

B. Real options policy as approximate dynamic programming method for network design and timing

- Monte-Carlo simulation (Boyle 1977)
  more effective than finite difference binominal lattice especially under multiple and complex uncertain elements, where uncertainties can be modeled as non-stationary stochastic processes
- LSMC: Least Squares Monte Carlo (Longstaff and Schwartz 2001)
  Multi option LSMC (Gamba 2003) for interrelated options are further proposed
- Network-based RO Model (Chow and Regan 2011a,b)
- Sampling technique to obtain an extreme value distribution (Chow and Sayarshad, 2016)
  ❌ Estimation is prone to high variability based on sample size
  ❌ Policy that remains to those values remains unknown

Related Literature 3/3

C. Neural networks for sequence understanding

- RNN (Hochreiter and Schmidhuber 1997; Rumelhart et al. 1986)
  Specifically designed for understanding sequences and significant success even for non-trivial sequential patterns
- LSTM (special type of RNN) for predicting ride-hailing demand (Jin et al. 2020)
- Transformer neural network (Vaswani et al. 2017)
  has parameters in the order of hundreds of millions, and require self-supervised learning based pre-training strategies, which is domain-dependent and not easy to figure out

This paper don’t employ transformer but RNN for its exploratory nature of this work since RNNs have far fewer parameters to learn and pre-training is not necessarily required to learn non-trivial patterns

Demand specification

At each decision epoch $t_d \in T$,
the aim is to determine the investment/deferral strategy $a_h : \{1, 0\}$
for each $h \in H_{cand}$ s.t. $H_{cand} \subseteq H$

$T = \{t_1, t_2, \dots, t_E\}$: The set of discrete time periods

H: Set of candidate zones

H_{sub}: Subzone of $h \in H$; $H_{sub} \times H_{sub}$ represents OD pairs

$Q_{ijt} \in \mathbb{R}^{|H_{sub}| \times |H_{sub}|}$: OD demand for the MoD service at time $t_n \in T$

- OD pair ij can be characterized by a stochastic process such as Geometric Brownian motion (GBM). <- related to search theory
- This is assumed to be independent between zones though correlated multivariate process can be considered with appropriate data.

Demand specification

Service demand at each OD pair ij satisfies GBM form

Stochastic process Q_{ijt} follows the stochastic differential equation below

$$\frac{dQ_{ijt}}{Q_{ijt}} = \mu Q_{ijt} dt + \sigma Q_{ijt} dW_t \quad (2)$$

dt: infinitesimal time increment
μ: drift parameter (revenue rate)
σ: volatility rate of service demand
σ_H: zone-specific volatility is assumed (not for individual heterogeneity)
$dW_t \sim N(0, dt)$: Standard Wiener process

Investment payoff

Aggregated network model for urban taxi service to estimate OD ridership for MoD service for hourly MoD ridership λ_{ij}

$$\lambda_{ij} = Q_{ij,t_n} \exp^{-\gamma(c_{ij} + \alpha_{IV}.VoT.TIV_{ij} + \alpha_W.TW)} \quad (3)$$

VoT: Value of time of MoD users in service region
TIV_{ij}: In-vehicle travel time between OD ij
c_{ij}: Trip price for OD ij

TW: Expected wait time
γ: Congestion scaling parameter $[0,1]$
α_{IV}, α_W: Customer perception factors for in-vehicle time and wait time

Under the assumption that the taxi dispatching market where the number of vehicle is optimum, following is the estimated wait time (Salanova et al. 2014)

$$TW = 0.8\lambda_u^{1/3} v^{-2/3} \quad (4)$$

v: vehicle speed / λ_u: hourly MoD ridership in a region given by Eq.(5)

$$\lambda_u = \sum_i \sum_j \lambda_{ij}, \ i \in \mathscr{H}_{sub}, j \in \mathscr{H}_{sub} \quad (5)$$

Investment payoff

Cumulative (peak hour) MoD ridership is calculated using Alg. 1

Algorithm 1 MoD ridership calculation for a subset of Z service zones in a region with $\mathscr{Z}_{sub}$ sub-zones

```
Initialize λ_{0,ij} = Q_{ij,t_n}; i, j ∈ 𝒵_sub
Calculate λ_{0,u} = f(λ_{0,ij}); i, j ∈ 𝒵_sub using Equation (5)
Calculate TW_0 = f(λ_{0,u}, v) using Equation (4)
Initialize gap_wait = 1000 (min), itr = 0, tol = 0.001 (min)
while gap_wait ≥ tol do
    Update itr = itr + 1
    Calculate λ_{itr,ij} = f(λ_{itr−1,ij}, TW_{itr−1}, c_ij, VoT, TIV_ij, γ, α_IV, α_W); i, j ∈ 𝒵_sub using Equation (3)
    Calculate λ_{itr,u} = f(λ_{itr,ij}); i, j ∈ 𝒵_sub using Equation (5)
    Calculate TW_itr = f(λ_{itr,u}, v) using Equation (4)
    Update gap_wait = TW_it − TW_{itr−1}
Cumulative ridership for Z zones = λ_{itr,u}; OD ridership for 𝒵_sub sub-zones in Z zones = λ_{itr,ij}
```

Investment sequence s: $\{z_1, z_2, \ldots, z_H\}$,

Cumulative ridership from addition of z_h in a region with $z_1, \ldots, z_{h-1}$ is returned as $X_{z_h}^{cum}$ using Alg. 1

Service zone ridership for z_h can be calculated as $X_{z_h}^{cum} - X_{z_{h-1}}^{cum}$

Investment payoff

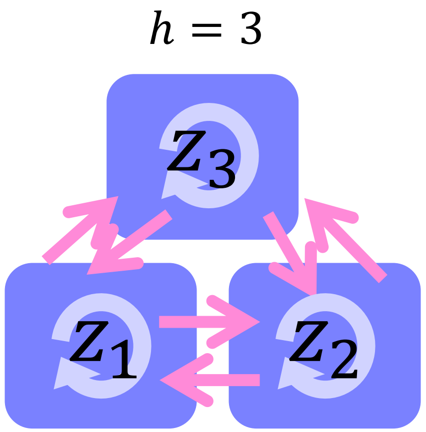
- The investment payoff for zone z_h is ridership gained from including zone z_h minus a ridership threshold.

$$\pi_{z_h} = X_{z_h} - (C_{wz} + 2(h-1)C_{iz}) \quad (6)$$

C_{wz}: within-zone cost for within-zone riders / C_{iz}: inter-zone cost for inter-zone riders

- $2(h-1)$ new connections with preceding h-1 zones in the sequence become open by increasing one zone

$\{z_1, z_2, z_3\}$: a sequence for a region with three zones



Additional 1 within-zone link and $2(h-1)$ inter-zone links are assumed

Solution algorithm

- S denotes a set constituting $H!$ sequences.

- To determine the optimal investment strategy subject to the compound options, LSMC method is adopted for estimating the policy value of each investment sequence $s \in S$ s.t. $|S| = L$

- For a given set of H zones and set of zone-specific volatility σ_H, MC simulation returns each of the $H_{sub} \times H_{sub}$ OD pair service demand by generating paths $\mathcal{P}$ by means of GBM for each $t_n \in T$.
  - Continuation value at t_n is then estimated given a realization of $\mathcal{P}$

Solution algorithm

The goal

Search for the sequence that gives the highest option value (Chow and Regan, 2011*a*) which forms the optimal investment strategy (*i.e.*, invest now or defer later for each zone $h \in H$)

HOW?

Estimate the option value for each of the enumerated sequences in S using CR policy which is VFA method in approximate dynamic programming (Powell, 2007) to solve the lower bound MoD service region design and timing problem with stochastic variables.

Solution algorithm

Challenge

There is increasingly high computational cost in determining the policy value for all $H!$ sequences for a large number of zones.

Main idea

ML based CR-RNN policy to accelerate the calculations for efficiently determining the optimal investment strategy for the service region design.

CR policy

- CR policy is categorized into two main components

SEQ: Enumeration of possible $H!$ permutations (ordered sequences in set S) for H zones

ROV: Real option valuation of each sequence in S. The ROV component majorly contributes to computation time in the RO model, and hence, is the focus of this study.

Overview of value estimation steps

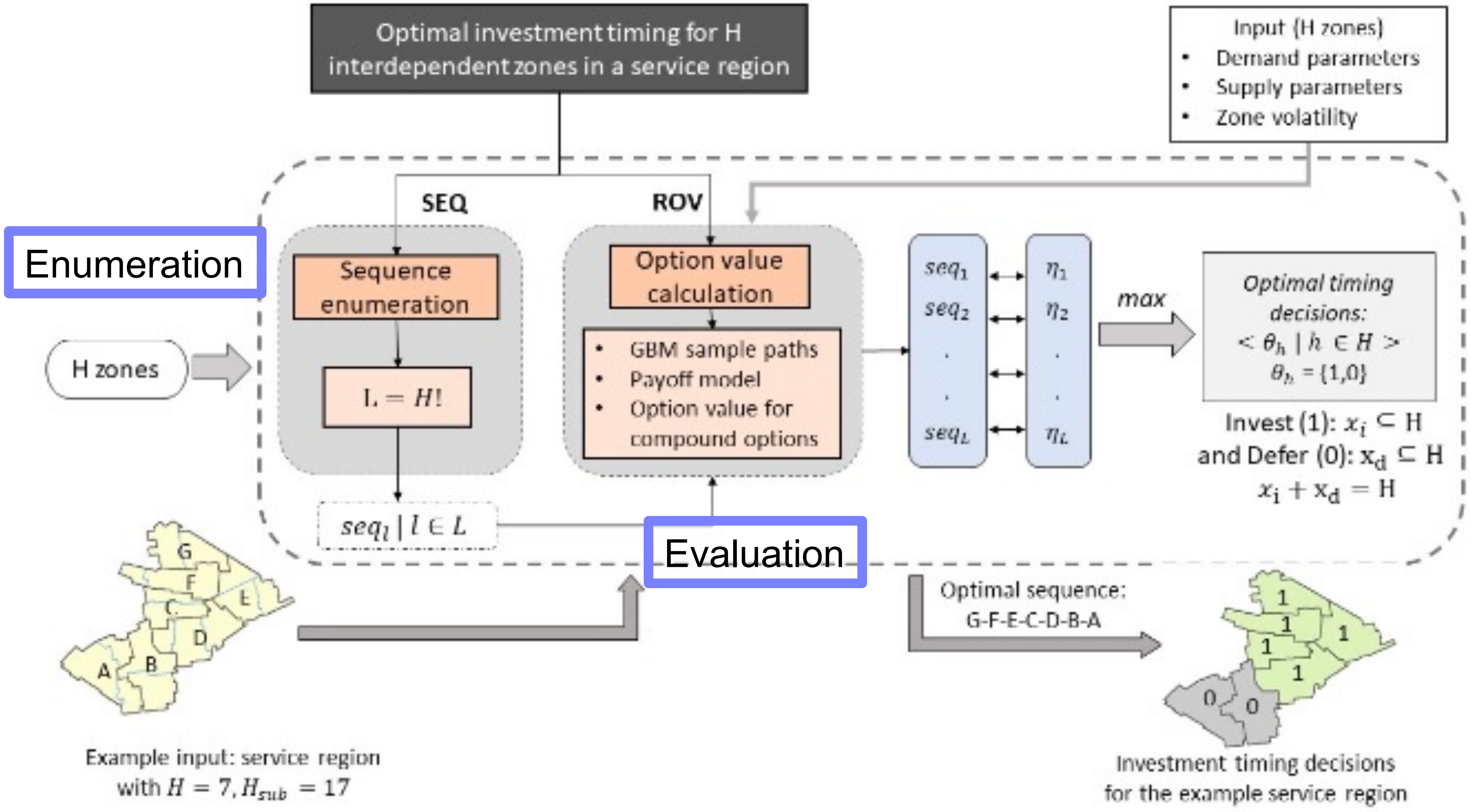


Fig. 3: CR-RNN policy model for determining which mobility service region zones to serve now and which to defer (zones and sub-zones illustrated in the figure are PUMA and taxi zones in NYC respectively).

ROV: option value calculation

- Consider an investment sequence $s \in S$ where $s = \{z_1, \dots, z_H\}$. The value of the subsequent options is incorporated when valuing the deferral option for any project $z_h \in s$.

Bellman Eq.

for determining the value of the option to invest in a service region design as a function of stochastic OD demand.

ρ: the risk-free discount rate

immediate investment of z_h — opportunity of investment of z_{h+1}

$$F_{z_h}(t_n, X_{z_h,t_n}) = \max\{\pi_{z_h}(t_n, X_{z_h,t_n}) + F_{z_{(h+1)}}(t_n, X_{z_h,t_n}), (1+\rho)^{-(t_{n+1}-t_n)} \mathbb{E}_{t_n}[F_{z_h}(t_{n+1}, X_{z_h,t_{n+1}})]\},$$

X_{z_h,t_n}: state variable of the project, that is, estimated MoD ridership in zone z_h at t_n

$\pi_{z_h}(t_n, X_{z_h,t_n})$: net present worth of an immediate investment of a project z_h at t_n

$\mathbb{E}[F_{z_h}(t_n, X_{z_h,t_n})]$: expectation function of all future investment benefits from the contingent claim optimally exercised at a certain time. (representing the continuation value $\phi_{z_h}(t_n, X_{z_h,t_n})$)

ROV: option value calculation

Eq.(8)-(10) show the continuation value calculation of service zone $z_h \in s$ at time step t_n for a sample path p

$$\phi_{z_h}(t_n, X_{z_h,t_n}(p)) = \mathbb{E}^*_{t_n}\left[\sum_{i=n+1}^{E}(1+\rho)^{-(t_i-t_n)}\sum_{r=h}^{H}\pi_{z_r}(t_n,t_i,\tau,p)\right] \quad (8)$$

$$\text{with } \pi_{z_h}(t_n,t_i,\tau,p) = \begin{cases}\pi_{z_h}(t_i, X_{z_h,t_i}(p), & \text{if } t_i = \tau_{z_h}(p)\\ 0, & \text{otherwise}\end{cases} \quad (9)$$

$$\Longrightarrow \phi_{z_h}(t_n, X_{z_h,t_n}(p)) \approx \phi^*_{z_h}(t_n, X_{z_h,t_n}(p)) = \sum_{i=1}^{J}\beta^*_j(t_n)L_j(X_{z_h,t_n}(p)) \quad (10)$$

β_J^* are the optimal coefficients for the J basis functions obtained using least-squares estimation

$\tau_{z_h}(p)$:The optimal stopping time for a zone z_h in the p-th path

$\theta_{z_h,t_n} \in \{0,1\}$: The decision to invest or exercise immediately in zone z_h at time t_n
(1 is invest now / 0 is defer for later)

L_j is Laguerre polynomials and can be calculated analytically. $L_n(x) = e^x \frac{d^n}{dx^n}(e^{-x}x^n)$

ROV: option value calculation

- If the payoff of z_h exceeds the continuation value at t_n, the investment is exercised and the optimal stopping time is updated to t_n; this is done recursively from maturity t_E to t_1

- Using the optimal stopping times of each MC path ($p \in P$), which includes the earliest investment timing of each zone $z_h \in s$, the option value at t_0 is determined

- The investment sequence that offers the highest option value for the initial project is selected.
  - η_s denotes policy value of sequence s. All η_s for total L sequences are calculated.

ROV: option value calculation

- Some of the sub-sequences in H! are repeated and the cumulative benefits (i.e. ridership and payoff) of the option z_h for a sequence s_1 at time t_n can be re-used for an option in another sequence s_2.
- The effective computational cost of evaluating $H!$ sequences is on the order of

$$O\left(\sum_{h=1}^{H} \boxed{\frac{H!}{H-h!}} \times |\mathcal{P}| \times |S| \times \text{payoff}\,(\text{H})\right)$$

overlapping repeated sequences is avoided for calculation

Challenge

✓ Incorporating endogenous nature of demand and investment decisions
… The above framework assumes exogenous demand and zone volatility

✓ Valuations of parameters regarding the investment payoff model
… This instead makes ROV calculations more computationally expensive

▶ CR-RNN policy as an efficient approach to obtain the optimal strategy

The objective

Efficiently obtain the invest sequence out of $H!$ sequences that offers the highest initial project value, which forms the optimal investment strategy.

HOW?

- Sampling M sequences ($|S_m| = M$) to train RNN that can identify a small set of promising sequences
- Labeling the sequences based on policy values, assuming that they follow Weibull distribution
  - 50 percentile of the Weibull CDF is selected as the threshold to label 1 or 0 for the sequence s
- Apply the RNN classifier on remaining sequences ($|S_r| = L - M$) to get the probability estimate of how likely the sequence is promising.
- ROV calculations are performed for K top-rated sequences ($|S_k| = K$) out of the remaining.
- By this way, ROV calculations are required only for $M + K$ sequences out of L sequences.

Illustration of the CR-RNN policy

- Randomly sample a small fraction of sequences ($frac_{seq}$) from the population set S

- S_m denote the selected subset of sequences such that $|S_m| = M$ ($M = frac_{seq} \times L$).

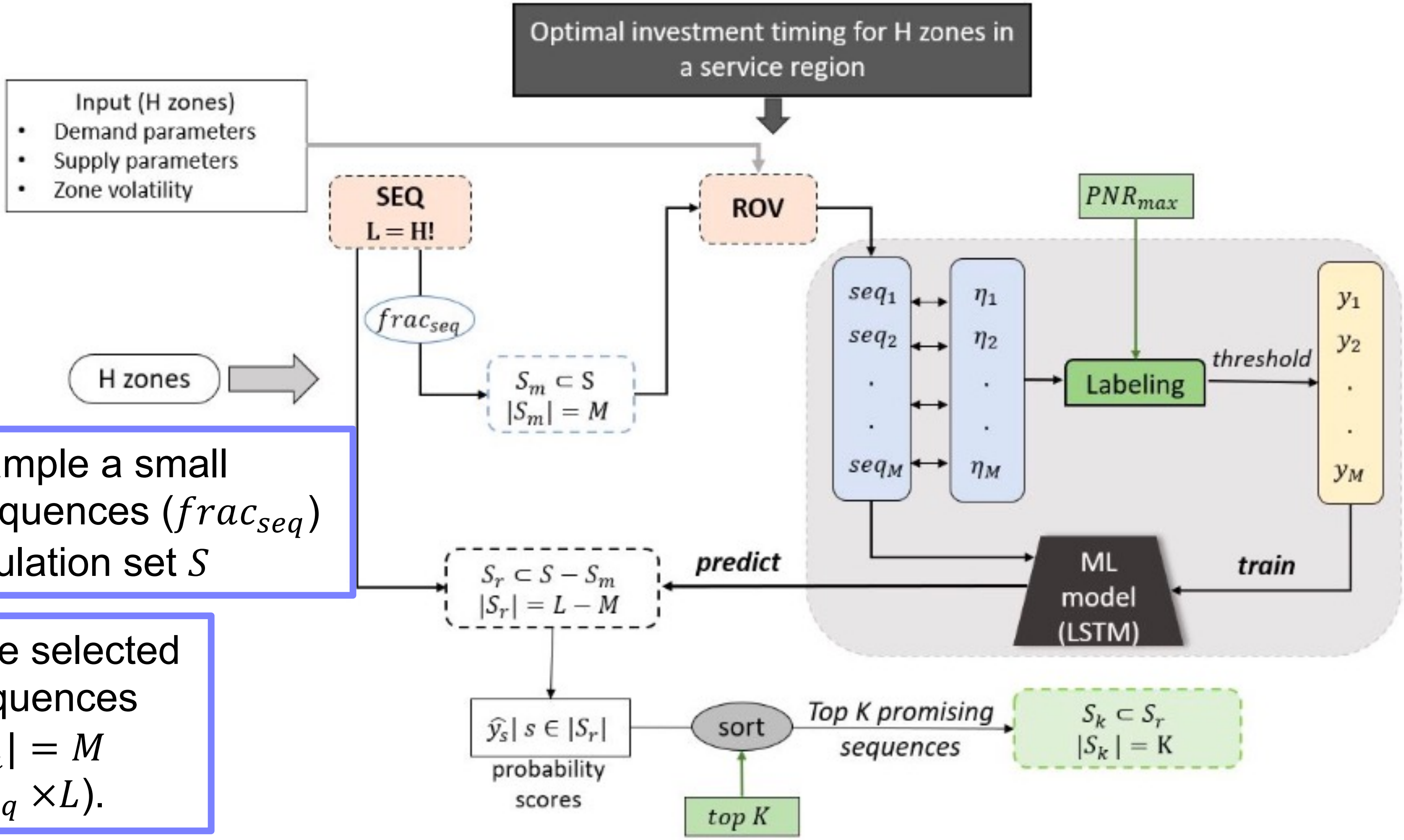


Fig. 4: Using the RNN based sequence classification in the CR-RNN policy for efficient ROV calculations

Classification of a candidate sequence s as a promising sequence

① Input embedding layer:
Embedding for z_h denoted by $emb(z_h)$ which contains features of zone z_h

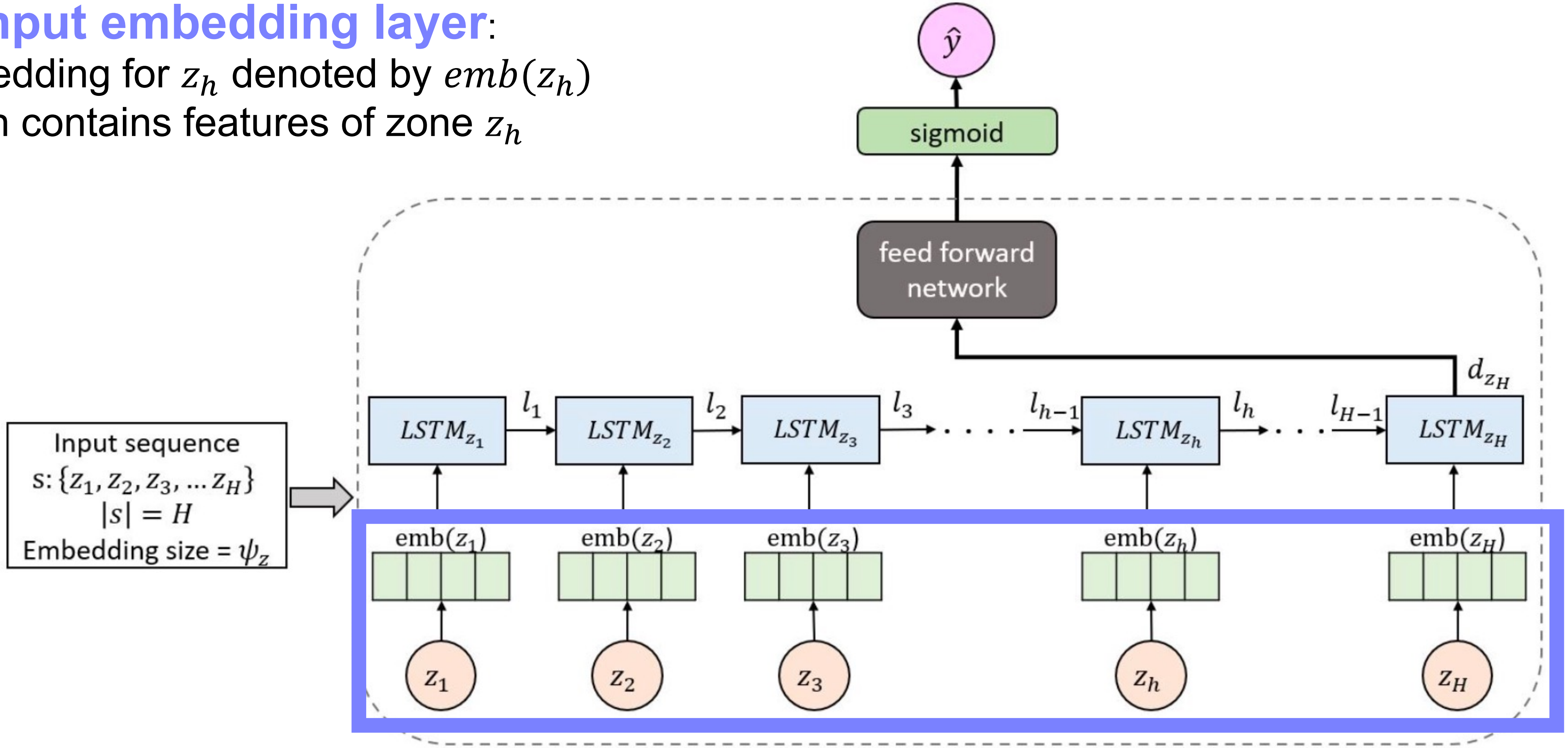


Fig. 5: LSTM based RNN architecture for investment sequence classification.

Classification of a candidate sequence s as a promising sequence

② LSTM layer:

- Hidden state of $LSTM_{z_h}$ depends on both $emb(z_h)$ and the previous unit $LSTM_{z_h-1}$ state.
- Standard update rules for LSTM state update, with input gate/output gate/forget gate included.
- Last unit is composed of hidden state (d_{z_H}) and cell state(c_{z_H})

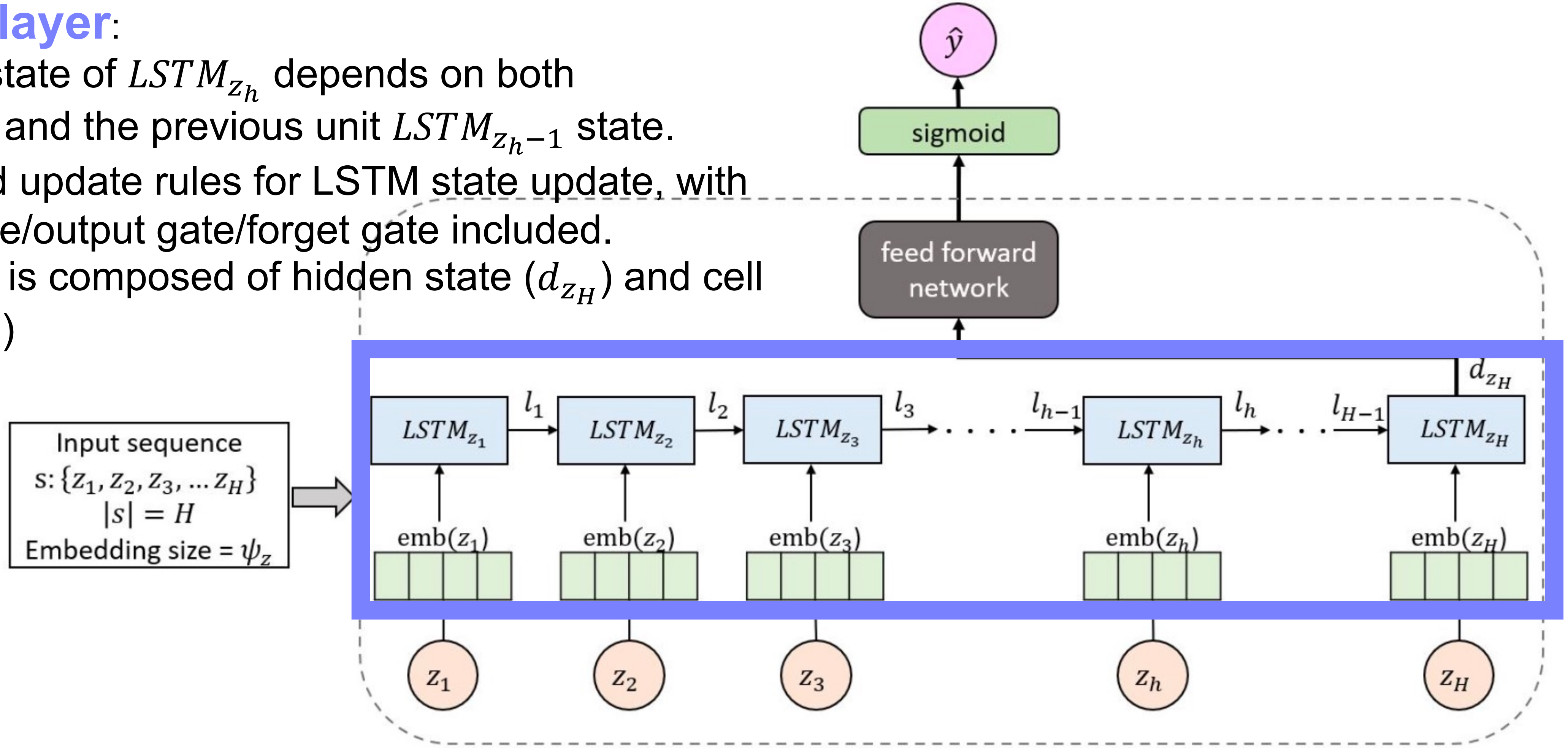


Fig. 5: LSTM based RNN architecture for investment sequence classification.

Classification of a candidate sequence s as a promising sequence

③ Feed forward layer:

- Hidden state (d_{z_H}) is fed to a feed forward neural network
- $l_{FF} = W_{FF} d_{z_H} + b_{FF}$
  where $W_{FF} \in \mathbb{R}^{1 \times \psi_z}$ is the weights matrix,
  $d_{z_H} \in \mathbb{R}^{\psi_z}$ is the final LSTM unit's hidden state,
  and $b_{FF} \in \mathbb{R}$ is the bias term
- $l_{FF} \in \mathbb{R}$

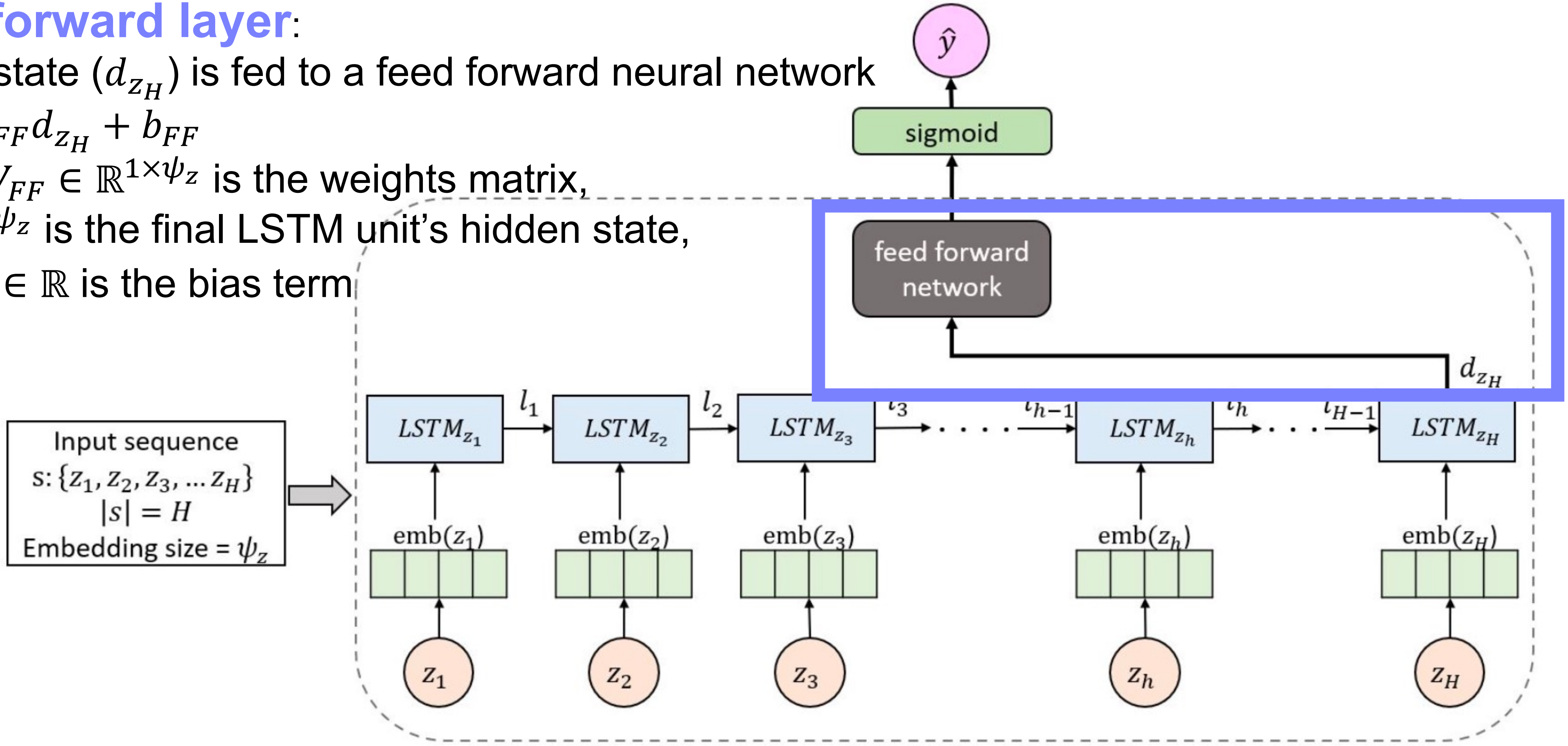


Fig. 5: LSTM based RNN architecture for investment sequence classification.

Classification of a candidate sequence s as a promising sequence

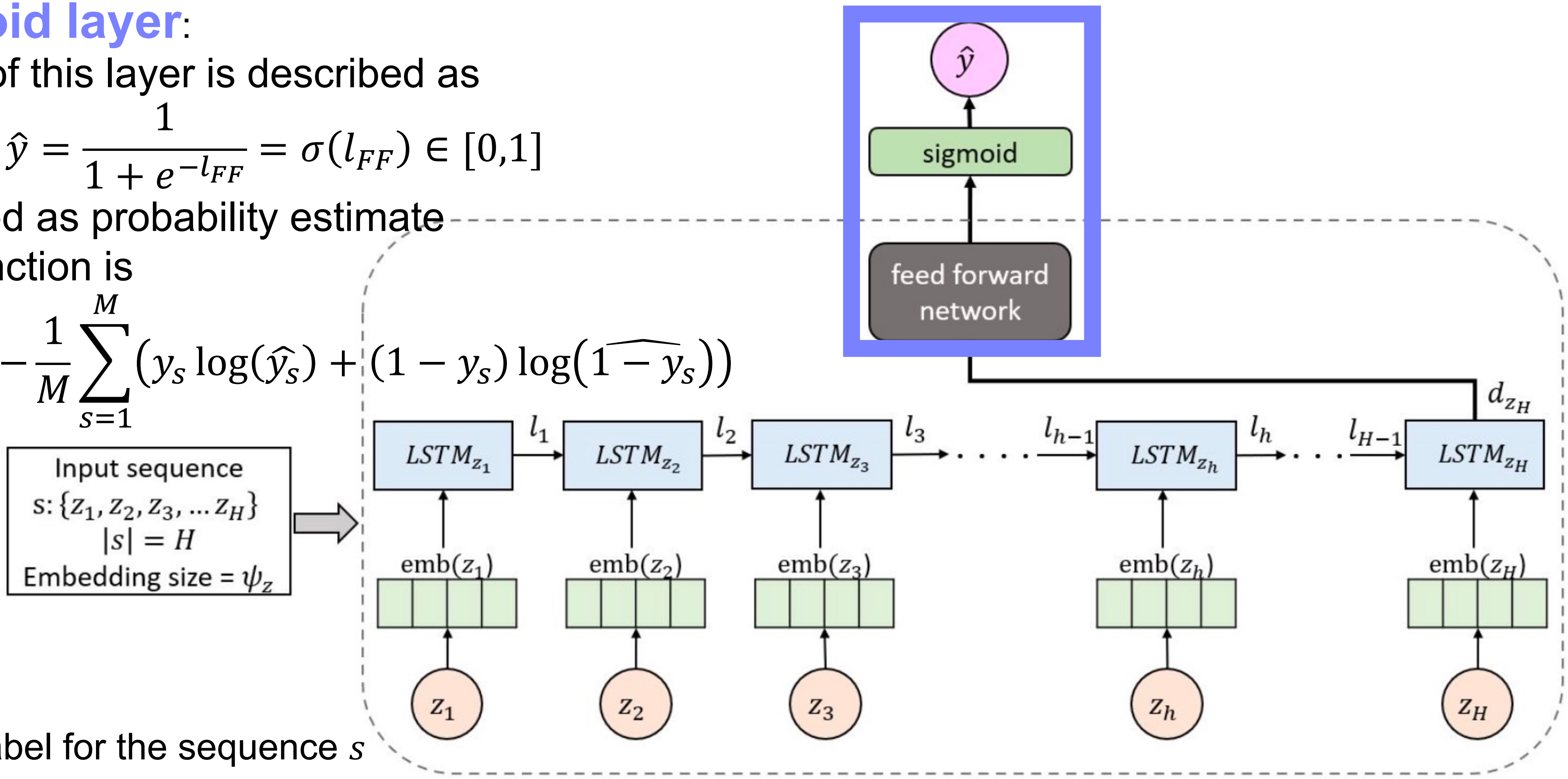
④ Sigmoid layer:

- Output of this layer is described as

$$\hat{y} = \frac{1}{1 + e^{-l_{FF}}} = \sigma(l_{FF}) \in [0,1]$$

- $\hat{y}$ is used as probability estimate
- Loss function is

$$loss_C = -\frac{1}{M}\sum_{s=1}^{M}\left(y_s \log(\hat{y_s}) + (1 - y_s)\log\left(\widehat{1 - y_s}\right)\right)$$



※ y_s is a binary label for the sequence s

Fig. 5: LSTM based RNN architecture for investment sequence classification.

RNN model evaluation

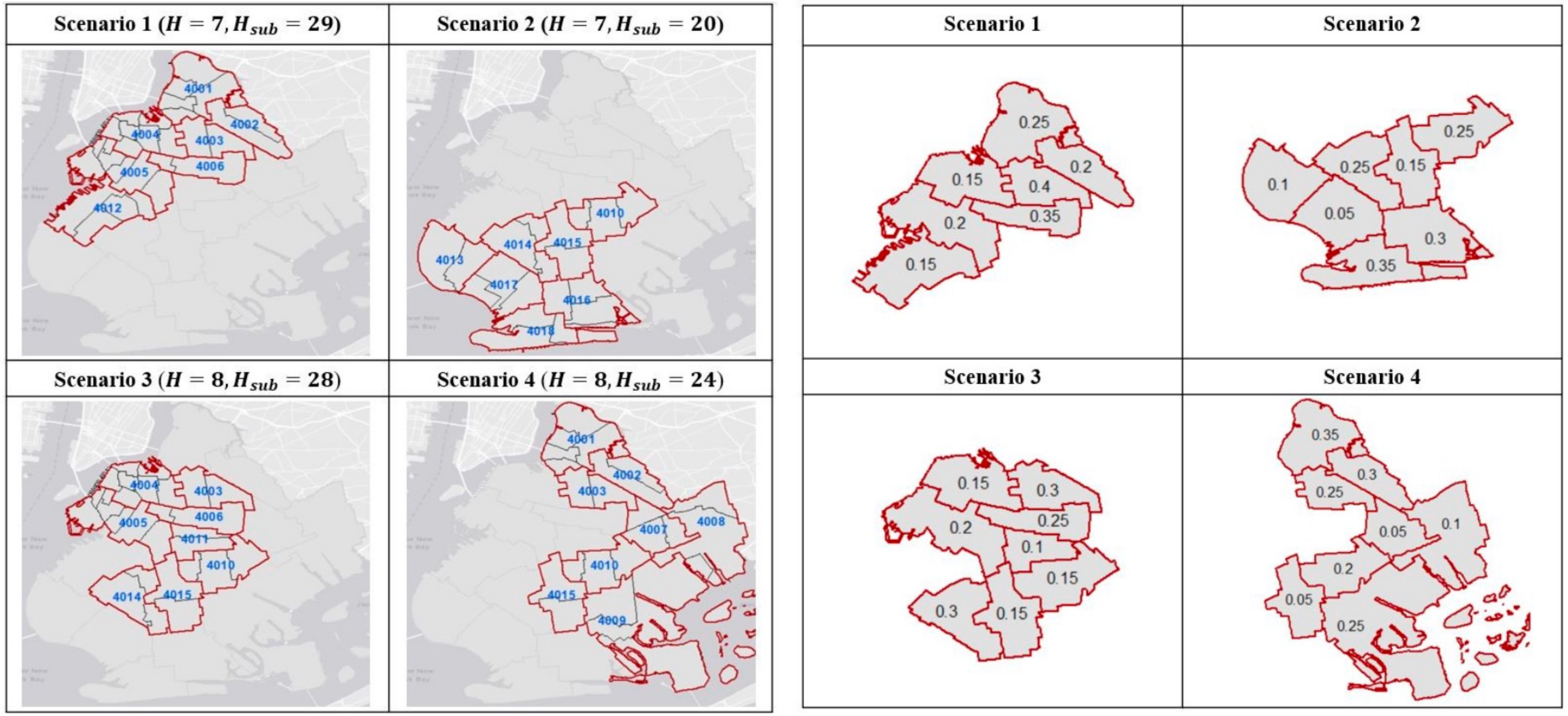
$$Gap@K = \frac{\eta_{true} - \eta_{pred}}{\eta_{true}} \times 100\%$$

η_{pred} : Policy value of the best among the top K sequences predicted by ML model
η_{true} : True value of the best sequence in $L - M$ test samples

The lower the $Gap@K$, the better is the performance of the model performance to varying input values

Experiment target area

- 4 different service region scenarios in Brooklyn NYC
  - OD service demand in $H_{sub} \times H_{sub}$ sub-zones evolve independently as GBM motions with zero drift and heterogeneous volatility



(a) (b)

Fig. 6: Service region scenarios in NYC: (a) shows PUMA zones (service zones) along with zone IDs and PUMA covered taxi zones (sub-zones) highlighted in red and black outline respectively, (b) shows the service zone volatility values (defined in Section IV-Ba).

Demand data specification

- Potential demand for MoD services is assumed to come form the auto and transit in the study are. Auto and public transit demand data is from Census Transportation Planning Products Program (CTPP, 2016).

- Of the aggregated commute flow across different taxi zones using, 23.38% was extracted as peak hour commute flow, and 60% of the above demand was defined as the potential OD demand for MoD services in the experiment.

Demand data specification

a) Parameters for ROV calculations in CR-RNN policy: We assume the following for the ROV calculations.

- Maturity time (t_E) = 5 years; $\mathscr{T}$ = {1,2,3,4,5}
- Number of basis functions (J) = 3
- Discount rate (ρ) = 2%
- Number of simulation paths ($|\mathscr{P}|$) = 300
- Value of time (VoT) = 0.293 euro/min (assuming \$20/hr as per prior studies (Holguín-Veras et al., 2012; Ma and Chow, 2022))
- Customer perceived wait time factor (α_W) = 2.1 (Kittelson & Associates, 2003)
- Customer perceived in-vehicle time factor (α_{TIV}) = 1 (Kittelson & Associates, 2003)
- Congestion scaling γ = 0.005 (based on Wong et al., 2001)
- Vehicle speed (v) = 19.31 km/hr (assuming 12 miles/hr average taxi speed in NYC (Liu, Vergara-Cobos and Zhou, 2019))
- Trip cost ($c_{ij}, i, j \in \mathscr{H}_{sub}$) = 2.42 euros (assuming a fixed price of \$2.75[8])
- Zone volatility values (Z_{vol}): {5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%}
  For each scenario with H zones, these values were randomly assigned
- Interzone cost (C_{iz}): The average interzone ridership is used to define the interzone cost for service zones
- Within-zone cost (C_{wz}): The ridership threshold that defines the within-zone cost for a set of service zones is considered to be 40% of the average within-zone ridership

Demand data specification

b) Parameters for RNN model: : We consider the following input parameters for the RNN model

- Training sample ratio ($frac_{seq}$): {0.0012, 0.0028, 0.0044, 0.0068, 0.0084, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5}, where $frac_{seq}$ = 0.01 means 1% of total population sequences (L) is used to train the RNN model.
- Top K sequences (K): {30, 50, 70}
- Embedding sizes (ψ_z): {10, 50, 100, 150, 200}
- Positive to negative (training) sample ratio (PNR_{max}) = {0.01, 0.02, 0.05}, where PNR_{max} = 0.01 denote 1% positive to negative samples in the M training set

- 20% of S_m sequences are set as a validation set for tuning hyper-parameters including epochs (up to 300) and embedding size from the list above.
- The detail of embedding is not written in the paper.

Optimal investment sequences

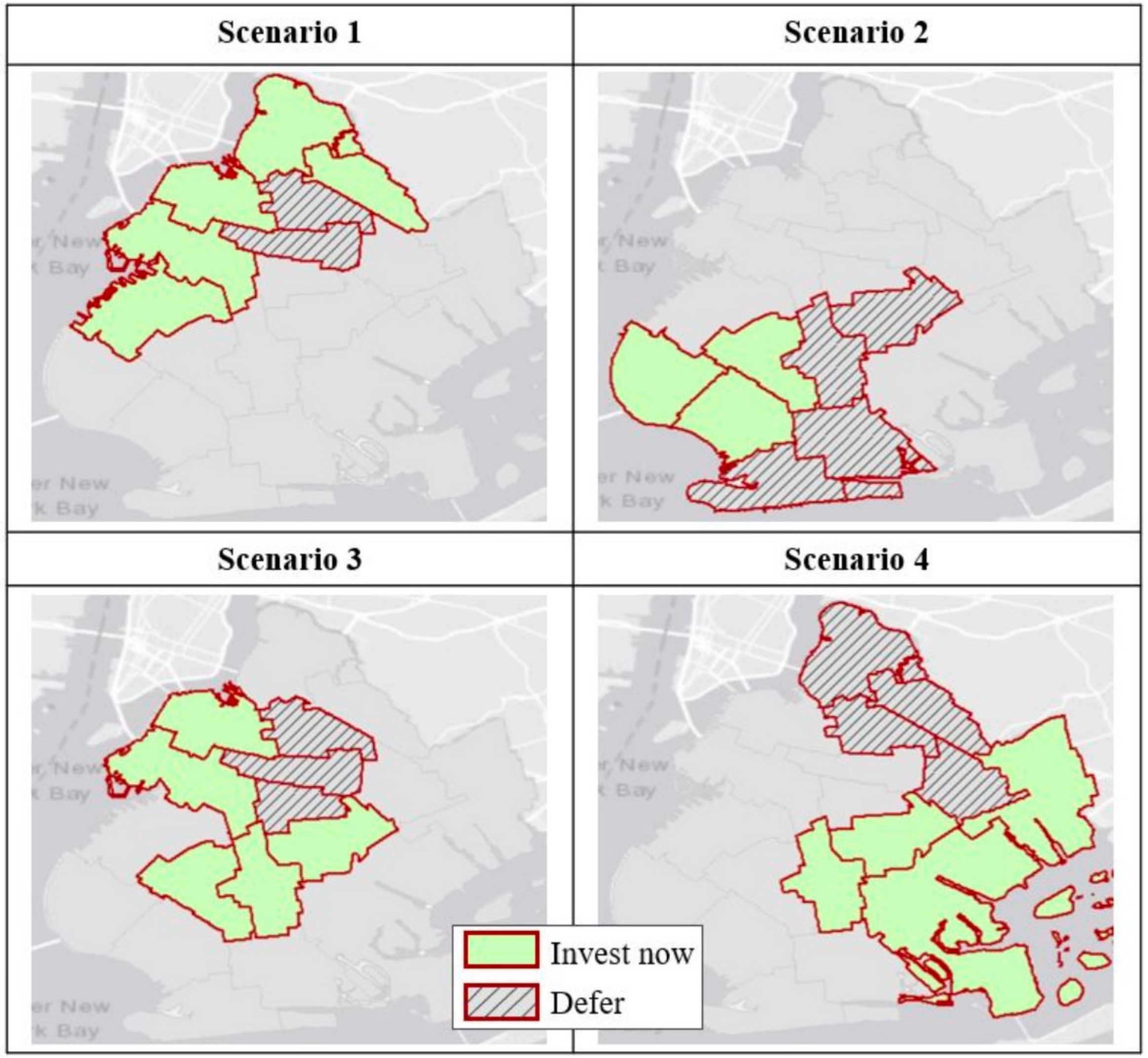


Fig. 7: Optimal investment strategy for service zones (*i.e.*, invest immediately or defer investment) obtained based on the CR policy for the service region scenarios.

Comparison between scenarios of CR policy

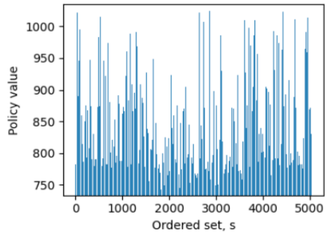
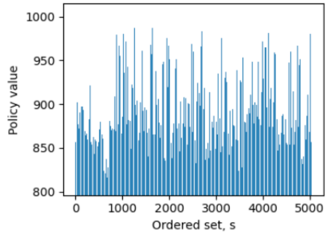
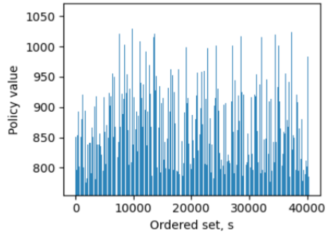
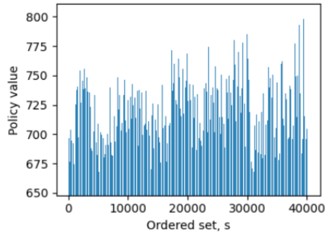
Scenario	C_{wz}	C_{iz}	NPV	Sequence with highest option value	Option value (highest)	Sequence with lowest option value	Option value (lowest)	Run time	Sequence policy value distribution
1	49	68	516	{4001, 4002, 4005, 4004, 4012, 4006, 4003}	1025.17	{4006, 4003, 4004, 4012, 4002, 4001, 4005}	738.4	0.69 hr	
2	65	59	684	{4017, 4013, 4014, 4016, 4018, 4010, 4015}	1005.05	{4016, 4018, 4015, 4017, 4010, 4013, 4014}	800.97	0.77 hr	
3	51	61	613	{4010, 4005, 4004, 4015, 4014, 4003, 4006, 4011}	1061.19	{4003, 4014, 4010, 4015, 4005, 4011, 4006, 4004}	759.86	4.17 hrs	
4	45	44	538	{4010, 4008, 4009, 4015, 4003, 4007, 4002, 4001}	801.07	{4002, 4001, 4003, 4007, 4015, 4008, 4010, 4009}	652.88	4.32 hrs	

Table I: Scenario-wise input parameters and CR policy values.

Comparison between scenarios of CR policy

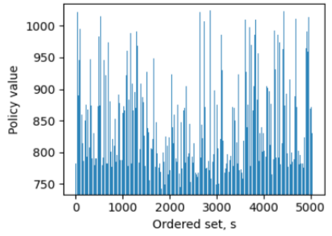
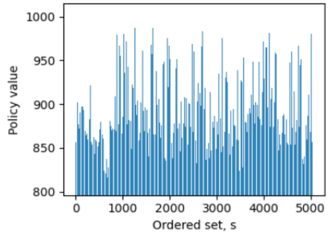
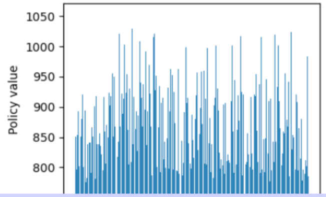
Scenario	C_{wz}	C_{iz}	NPV	Sequence with highest option value	Option value (highest)	Sequence with lowest option value	Option value (lowest)	Run time	Sequence policy value distribution
1	49	68	516	{4001, 4002, 4005, 4004, 4012, 4006, 4003}	1025.17	{4006, 4003, 4004, 4012, 4002, 4001, 4005}	738.4	0.69 hr	
2	65	59	684	{4017, 4013, 4014, 4016, 4018, 4010, 4015}	1005.05	{4016, 4018, 4015, 4017, 4010, 4013, 4014}	800.97	0.77 hr	
3	51	61	613	{4010, 4005, 4004, 4015, 4014, 4003	1061.19	{4003, 4014, 4010, 4015, 4005, 4011	759.86	4.17 hrs	

Table 1: Scenario-wise input parameters and CR policy values.

- Compared to investing in all zones immediately based on NPV, there is value waiting to invest in some zones using RO approach
- Lowest option values are 18-28 % lower that the highest option values for 4 cases, indicating that zone ordering can have the impact.

Comparison between scenarios of CR policy

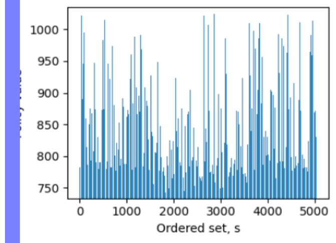
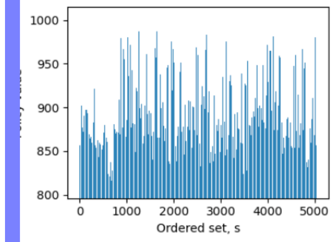
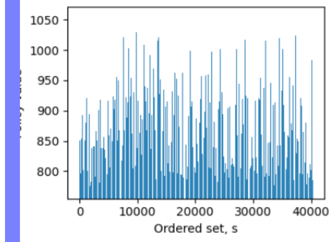
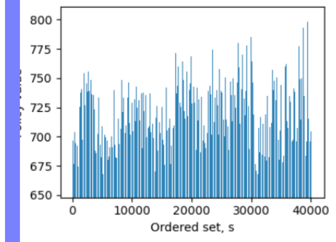
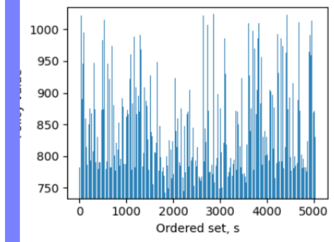
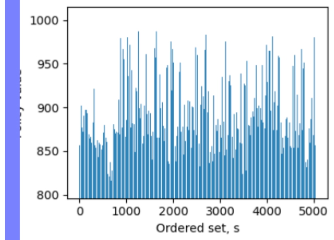
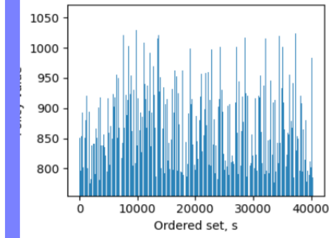
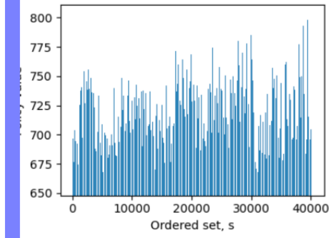
Scenario	C_{wz}	C_{iz}	NPV	Sequence with highest option value	Option value (highest)	Sequence with lowest option value	Option value (lowest)	Run time	Sequence policy value distribution
1	49	68	516	{4001, 4002, 4005, 4004, 4012, 4006, 4003}	1025.17	{4006, 4003, 4004, 4012, 4002, 4001, 4005}	738.4	0.69 hr	
2	65	59	684	{4017, 4013, 4014, 4016, 4018, 4010, 4015}	1005.05	{4016, 4018, 4015, 4017, 4010, 4013, 4014}	800.97	0.77 hr	
3	51	61	613	{4010, 4005, 4004, 4015, 4014, 4003, 4006, 4011}	1061.19	{4003, 4014, 4010, 4015, 4005, 4011, 4006, 4004}	759.86	4.17 hrs	
4	45	44	538	{4010, 4008, 4009, 4015, 4003, 4007, 4002, 4001}	801.07	{4002, 4001, 4003, 4007, 4015, 4008, 4010, 4009}	652.88	4.32 hrs	

Table I: Scenario-wise input parameters and CR policy values.

Comparison between scenarios of CR policy

- When the number of service region increases from 7 to 8, the average computation time increase about 6 times, holding the benefit of ML approach to reduce the burden.

...tion ...lue ...west)	Run time		Sequence policy value distribution
8.4	0.69 hr		
0.97	0.77 hr		
9.86	4.17 hrs		
2.88	4.32 hrs		

4002, 4001}   4010, 4009}

Table I: Scenario-wise input parameters and CR policy values.

RNN-CR policy

H=7 case

Fig. 8: Scenario 1 (7 zones) RNN classification model prediction gap ($Gap@K$) for different training ratios for each K in top K (in legend) with (a) PNR_{max} = 1%, (b) PNR_{max} = 2%, (c) PNR_{max} = 5%.

RNN-CR policy

H=8 case

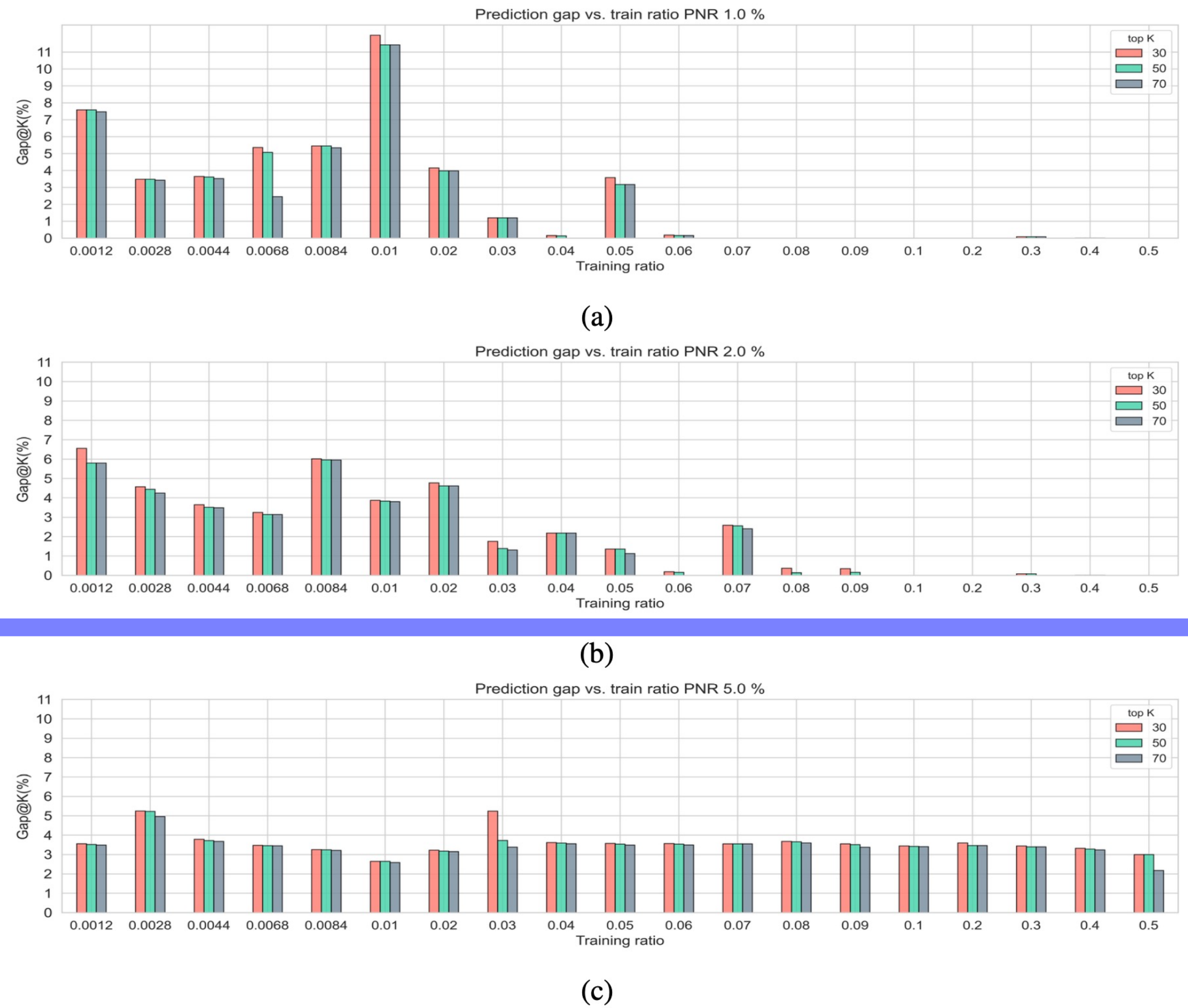
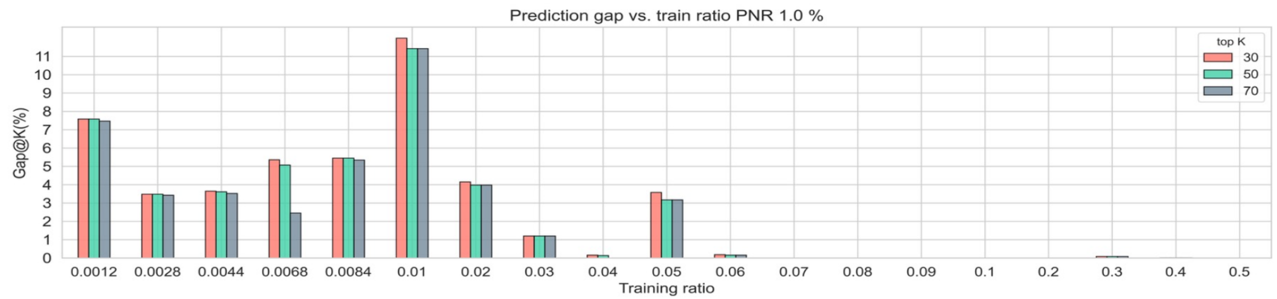
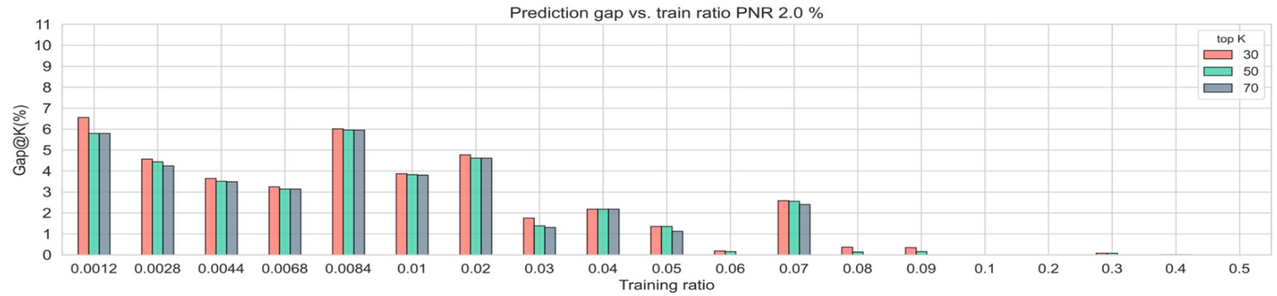


Fig. 10: Scenario 3 (8 zones) RNN classification model prediction gap ($Gap@K$) for different training ratios for each K in top K (in legend) with (a) PNR_{max} = 1%, (b) PNR_{max} = 2%, (c) PNR_{max} = 5%.

RNN-CR policy



(a)



(b)

- For PNR_{max} 0.01 and 0.02, the prediction gap monotonically decreases when training samples are increased.

Fig. 10: Scenario 3 (8 zones) RNN classification model prediction gap ($Gap@K$) for different training ratios for each K in top K (in legend) with (a) PNR_{max} = 1%, (b) PNR_{max} = 2%, (c) PNR_{max} = 5%.

RNN-CR policy

- For PNR_{max} 0.05, however, the prediction gap get worse
- This is because the number of positive samples increases due to a lower policy threshold value and can provide bad examples of promising sequences to the model

(a)

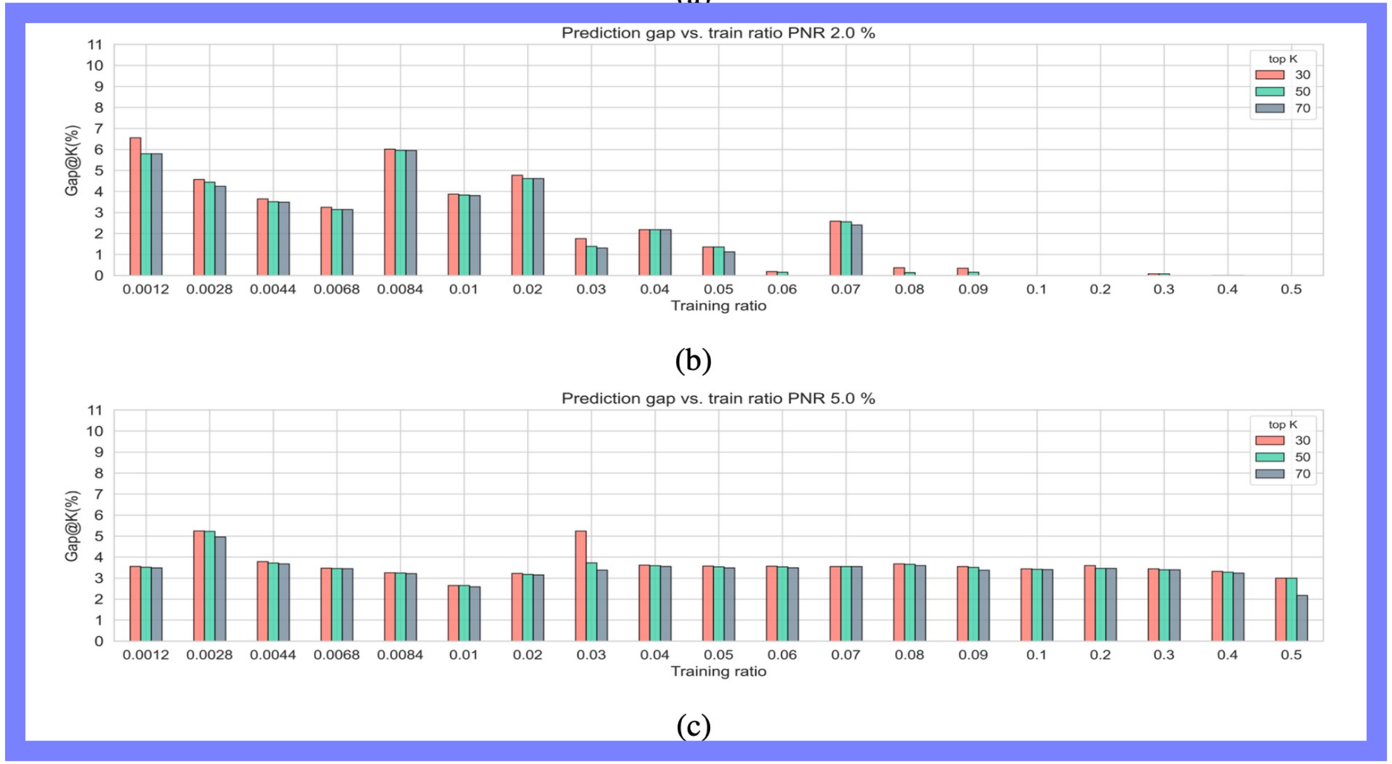


Fig. 10: Scenario 3 (8 zones) RNN classification model prediction gap ($Gap@K$) for different training ratios for each K in top K (in legend) with (a) PNR_{max} = 1%, (b) PNR_{max} = 2%, (c) PNR_{max} = 5%.

RNN-CR policy

Service zones	Training ratio (%)	0.12%	0.28%	0.44%	0.68%	0.84%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
7 zones	train time	2.42	2.15	2.35	2.76	2.58	2.26	3.32	3.84	2.89	5.16	5.11	5.5	6.87	9.2	7.63
	test time	0.56	0.55	0.62	0.6	0.6	0.56	0.58	0.55	0.67	0.66	0.62	0.61	0.57	0.56	0.61
	AUC	0.71	0.67	0.8	0.78	0.83	0.81	0.88	0.92	0.92	0.96	0.96	0.98	0.98	0.99	0.99
8 zones	train time	2.95	3.82	4.4	6.24	6.4	5.2	11.7	13.74	13.78	13.96	22.22	20.56	22.24	26.52	26.94
	test time	2.93	2.82	3.14	3.0	3.36	2.87	2.86	2.69	2.74	2.9	3.0	2.64	3.0	2.9	2.48
	AUC	0.8	0.8	0.86	0.92	0.91	0.88	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99

Table II: Average RNN classification model training time and testing time (in seconds) and test set AUC scores for different training ratio ($frac_{seq}$ in %) with PNR_{max} = 1%.

- AUC (Area under the curve) ROC (Receiver Operating Characteristics) curve (Murphy, 2012).
  - The AUC score (AUC $\in [0,1]$) tells us how well the model is able to distinguish between positive and negative classes.
- Achieved a substantial reduction in the overall computational cost

Summary

- Investment of Mobility on Demand (MoD) under demand uncertainty was translated into Real Options strategies which can incorporate the value of pending decision-making until the supplier get more information about demand or environment.
- LSMC simulation to evaluate the multiple interacting real options.
- For large-scale cases, proposing a supervised ML approach (using an RNN model) to address the high computational cost involved in valuating all possible $H!$ investment sequences via the CR policy.
- Train RNN classifier using the fractions of samples and determine best-K sequences from the rest of sampled sequences.

メモ(in Japanese)

リアルオプション

- 逐次最適化する事業計画を策定するにあたって不可欠となる視点は，「予算制約下において，どのタイミングでどれだけの事業投資を行うか」
- 投資のタイミングや金額の自由度，事業の中断・再開あるいは撤退の選択可能性といった意思決定の柔軟性を持つことは，より大きな収益を獲得できる可能性を生む，あるいは大きな損失を被る可能性を摘むことが出来るため，その投資事業に追加的な価値を付与する．これを「オプション価値」といい，事業価値評価においてオプション価値を勘案する考え方をリアルオプションという．

動的計画

動的計画：最終期以降の情報を利用して，本来の無限期間の問題を解くための枠組み
不確実性のある将来についてもその期待値を考慮して明示的に意思決定に組み込む

- 有限期間問題の最適値（費用）にその後に発生する期待費用を加えたものを最小化することを考える．無限期間の費用の期待値は無限大に発散するので，通常は将来の費用を適当な率で割引して期待値をとるか，確率1で到達する終端状態が存在すると仮定してモデル化を行う．
- 最終期末のシステム状態を入れるとその後の費用の期待値を返してくれる関数を到達費用関数(cost-to-go function)と呼ぶ．この関数が既知であるとすると，動的な無限期間の問題は，静的な有限期間の問題に帰着することができるが，実際には状態の数が膨大になり，各状態に対する期待値の計算も難しいので，このようなアプローチは不可能である．実際的なアプローチとしては，状態をより小さな空間に写像した特徴ベクトルとして保管し，特徴ベクトルに対する到達費用関数を，シミュレーションを用いて近似的に計算することによって最適化を行う．特徴ベクトルは対象とする問題の構造を熟慮して決められる．生産計画の例では，（一例ではあるが）各製品に対する在庫量が特徴ベクトルであり，輸送計画の例では，地域別の輸送手段の数などが特徴ベクトルの候補となる．

Rolling horizon method

無限期間で動的なモデルと，時間を止めた有限期間の静的なモデルの間には大きなギャップ
時間軸を持つ最適化モデルを考える際には不確実性情報の取り扱いが重要

- 不確実な未来の情報はわからないので，考慮の範囲外として既知の情報でモデルを構築するのが有限期間静的モデル．ここに時間が進むにつれて未来の不確定情報が徐々に確定していく様子を組み込んだのがローリングホライズン方式
  - 4週間分の生産計画を作成し，それを求解することによって得られた解にしたがって，最初の1週間の生産を行う．
  - 1週間経過した後は，すでに過去になった1週間分は固定し，その先の4週間分の生産計画を作成し直す．これは開始時から数えると5週間後までの計画を作成していることに相当する．

最適停止問題

- ROは最適停止問題として見ることができ，動的計画法で解ける

- 例．アメリカンコールオプションの定式化(タイミングの数理 p9より)
  金融商品のオプションの中で満期 T までの任意の時刻で権利行使可能なアメリカンコールオプションを考える．権利行使が発生する時刻を t として，初期原資産の価格を $X(t) = x$ とする．資産価値の確立過程 $\{X(t)\}$ が幾何ブラウン運動(対数Wiener過程)に従う，すなわち $\{X(t)\}$ が
  $$dX(t) = \mu X(t)dt + \sigma X(t)dB(t)$$
  で $B(t)$ はブラウン運動に従うとする．権利行使価格を K としたとき，オプションを時刻τで行使した時の期待利得 Y_τ は，
  $$Y_\tau = \max\{X(\tau) - K, 0\}$$
  となる．したがって問題は $E(Y_\tau | X(t) = x)$ を最大にする $t \in [t, T]$ を求める最適停止問題となる．

サーチ理論と最適停止問題の関係性

- Stiglerモデル(1961) The Economics of information

購入主体は一定費用 c の下で分布 F から独立した価格の抽出を行う. N 番目までの最小価格のとる期待値と $N-1$ 番目までのものを比較すると, サーチの限界便益は正である. すなわちサーチ数が増えるほど最小価格期待値は小さくなる. この正の限界便益がサーチコスト(定数)を下回るまで抽出を継続することが最適である => 最適停止点

Stiglerモデルへの批判 : https://www.glocom.ac.jp/wp-content/uploads/2020/10/78.pdf

a) **決定ルールの逐次性の欠如**: -> McCall(1965)
買い手は価格分布を知っているので事前に有限回のサーチ数を決定するが, この決定ルールでは価格サーチによって得られた情報を利用しておらず, できるだけ安い価格で買おうとする立場としては合理的な仮定ではない.

b) **供給側の決定ルール仮定の欠如**: -> Salop&Stiglitz(1977)
Stiglerの分析は需要者である買い手の行動のみを分析しており, 供給者である売り手の分析がなされていない.

逐次的サーチ過程

- McCall

既存の情報に則って受容かサーチ継続かの選択を逐次的に展開する逐次的サーチ・モデルを提示し，ある臨界値(留保価格)を境にして，それ以下の観察値に対し受容を選択し，それ以上のそれに対しサーチ継続を選択すべしとする最適停止ルール（optimal stopping rule）を導いた.

逐次的サーチ戦略の例：WeitzmanモデルによるPandora問題

確率変数が同一の分布に従う場合には全時間の戦略に共通の留保価格が得られる．このような即時効用と期待将来効用との均等化によって留保価格が算定されるような状況では，留保価格は他のすべてのサーチ機会とは独立であり，当該の確率変数自体の性質のみに依存する.

逐次的サーチ過程と動的計画法

- 伊藤過程にしたがって変動する状態変数をサーチ対象とする動的経済の下で，無限大時間視野，連続時間が想定されるところで展開される逐次的サーチ過程が導く最適停止ルール
  - Brown幾何運動(geometric Brown motion)は伊藤過程の特殊ケース
  - ↑後続価値に関するBellman方程式に対して最も簡単な形を生む連続時間確率過程であるために採用される
    - 条件付き確率分布 $\phi(x'|x)$ が正の系列相関性を持つ=第一次確率優位(first-order stochastic dominance)の要件を満たすことを保証する
    - この条件下において，臨界値前後で停止と継続が一意に分割できる
  - 臨界値において停止価値と後続価値か均等化し，それぞれが接線を共有するという追加条件のもとで，臨界値の値そのものが特定化される
    - これを平滑張合わせ条件(smooth-pasting condition)といい，数学的要請ではなく経済学的・物理学的・生物学的要請から従う

Ref: https://core.ac.uk/download/pdf/71792194.pdf(in Japanese)