Method

ColdNAS: Search to Modulate for User Cold-Start Recommendation

Task

Advisor : Jia-Ling, Koh

Speaker : Yu-Zhi, Liu

Source : WWW '23

Date : 2024/1/19

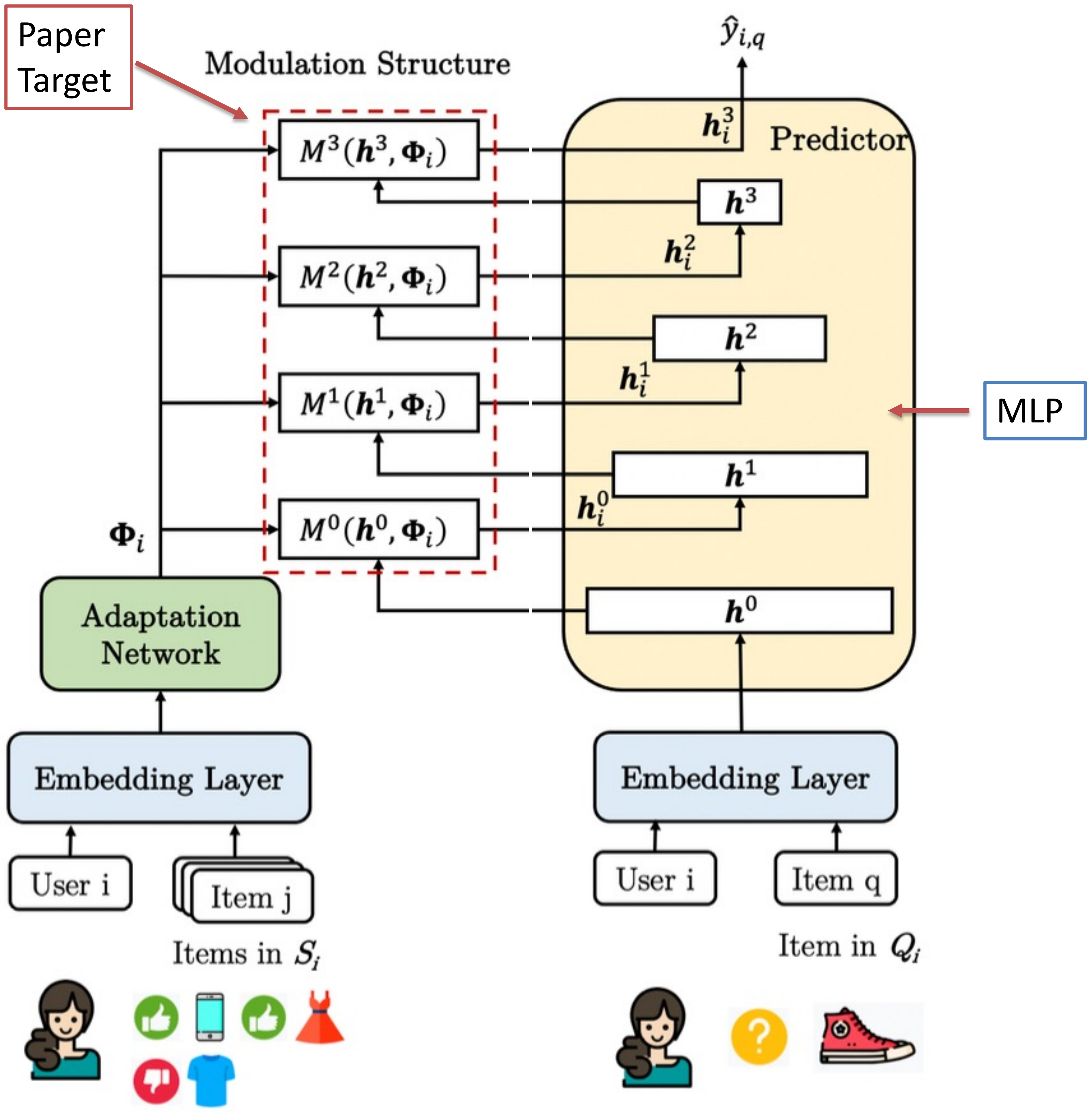
Outline

- Introduction
- Method
- Experience
- Conclusion

User Cold-Start Recommendation

- A recommendation system for cold-start users, who only have a few interaction histories.

Architecture



ColdNAS

Design

Search Space

Search Algorithm

Outline

- Introduction
- Method
- Experience
- Conclusion

Modulation-Based

Embedding layer E:

User & item features

$$(\boldsymbol{u}_i, \boldsymbol{v}_j) = E(u_i, v_j; \boldsymbol{\theta}_E).$$

Adaptation network A:

C is number of adaptive parameter groups

$$\boldsymbol{\Phi}_i = \{\boldsymbol{\phi}_i^k\}_{k=1}^{C} = A(\mathcal{S}_i; \boldsymbol{\theta}_A),$$

Predictor P:

Support set: $\mathcal{S}_i = \{(v_j, y_{i,j})\}_{j=1}^{N}$

$$\hat{y}_{i,j} = P((\boldsymbol{u}_i, \boldsymbol{v}_q), \boldsymbol{\Phi}_i; \boldsymbol{\theta}_P).$$

Multi-layer perception (MLP)

Madulation function

$$\boxed{\boldsymbol{h}^0 = (\boldsymbol{u}_i, \boldsymbol{v}_q)}$$

$$\boldsymbol{h}_i^l = M^l(\boldsymbol{h}^l, \Phi_i),$$

$$\boldsymbol{h}^{l+1} = \mathrm{ReLU}(\boldsymbol{W}_P^l \boldsymbol{h}_i^l + \boldsymbol{b}_P^l),$$

Learnable weights
at the Lth layer

Search Space

$$\boldsymbol{h}_i^l = \boldsymbol{h}^l \odot \boldsymbol{\phi}_i^1 + \boldsymbol{\phi}_i^2.$$

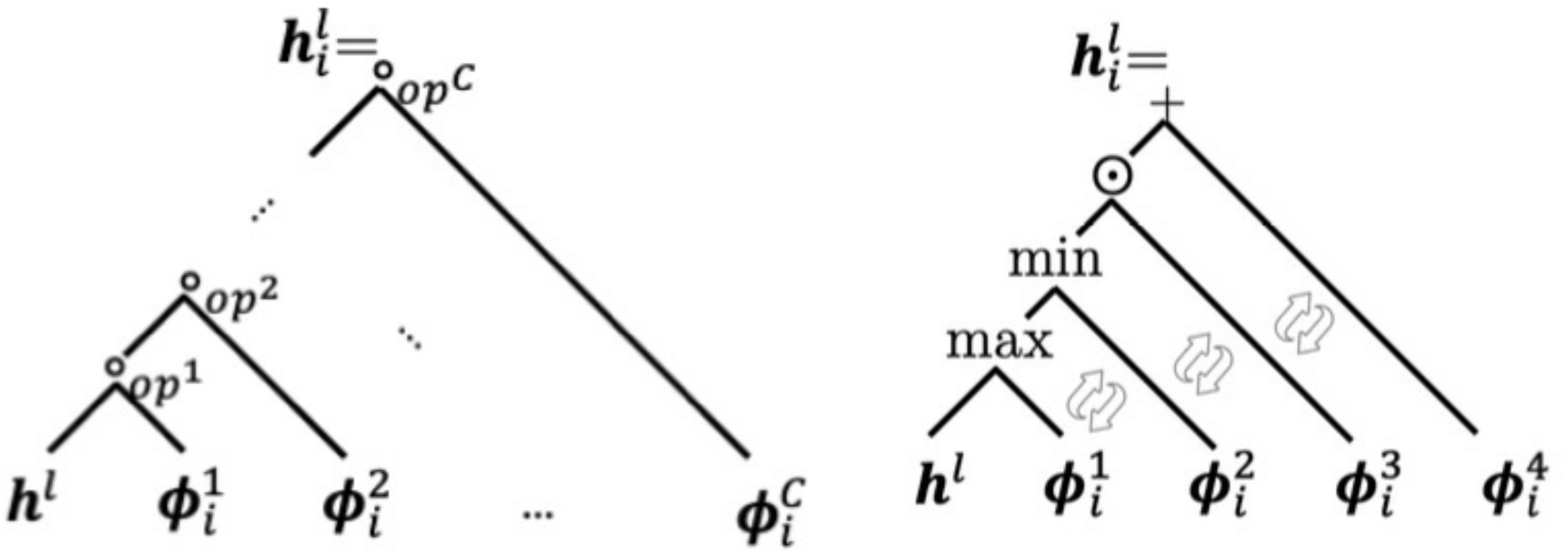
->FiLM

->ColdNAS

$$\boldsymbol{h}_i^l = \boldsymbol{h}^l \circ_{\text{op}^1} \boldsymbol{\phi}_i^1 \circ_{\text{op}^2} \boldsymbol{\phi}_i^2 \cdots \circ_{\text{op}^C} \boldsymbol{\phi}_i^C,$$

$$\circ_{\text{op}_i} \in \mathcal{O} \equiv \{\max, \min, \odot, /, +, -\}.$$

Search Strategy



(a) Original search space. (b) Transformed search space.

Original space size: $6^{C\times L}$. → Transformed space size: $2^{4\times L}$

Search Strategy Transformation

$$\boldsymbol{h}_i^l = \min(\max(\boldsymbol{h}^l,\ \hat{\boldsymbol{\phi}}_i^1),\ \hat{\boldsymbol{\phi}}_i^2) \odot \hat{\boldsymbol{\phi}}_i^3 + \hat{\boldsymbol{\phi}}_i^4, \tag{7}$$

(1) $\min(\max(\boldsymbol{h}^l,\ \boldsymbol{\phi}_i^1) + \boldsymbol{\phi}_i^2 - \boldsymbol{\phi}_i^3,\ \boldsymbol{\phi}_i^4) \odot \boldsymbol{\phi}_i^5$ *equals to* (7) *where* $\hat{\boldsymbol{\phi}}_i^1 = \boldsymbol{\phi}_i^1,\ \hat{\boldsymbol{\phi}}_i^2 = \boldsymbol{\phi}_i^4 - \boldsymbol{\phi}_i^2 + \boldsymbol{\phi}_i^3,\ \hat{\boldsymbol{\phi}}_i^3 = \boldsymbol{\phi}_i^5,\ \hat{\boldsymbol{\phi}}_i^4 = (\boldsymbol{\phi}_i^2 - \boldsymbol{\phi}_i^3) \odot \boldsymbol{\phi}_i^5$; *and*

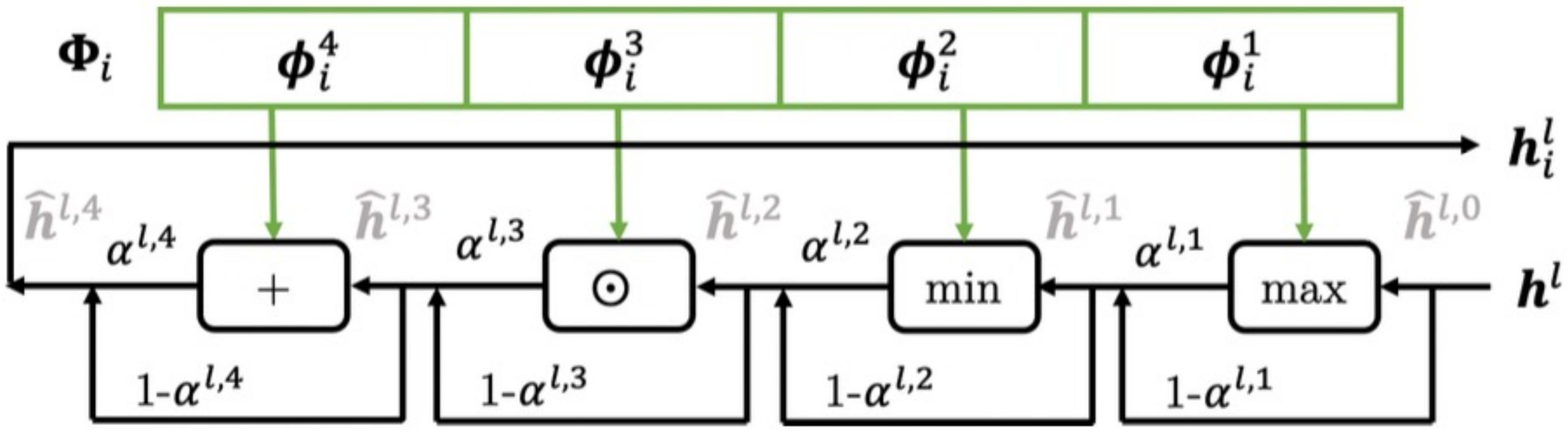
(2) $\max(\min(\boldsymbol{h}^l + \boldsymbol{\phi}_i^1,\ \boldsymbol{\phi}_i^2),\ \boldsymbol{\phi}_i^3) \odot \boldsymbol{\phi}_i^4$ *also equals to* (7) *where* $\hat{\boldsymbol{\phi}}_i^1 = \boldsymbol{\phi}_i^3 - \boldsymbol{\phi}_i^1,\ \hat{\boldsymbol{\phi}}_i^2 = \boldsymbol{\phi}_i^2 - \boldsymbol{\phi}_i^1,\ \hat{\boldsymbol{\phi}}_i^3 = \boldsymbol{\phi}_i^4,\ \hat{\boldsymbol{\phi}}_i^4 = \boldsymbol{\phi}_i^1 \odot \boldsymbol{\phi}_i^4$.

Construction of the Supernet

$$\hat{\boldsymbol{h}}^{l,k+1} = \alpha^{l,k+1}(\hat{\boldsymbol{h}}^{l,k} \circ_{\mathrm{op}^{k+1}} \boldsymbol{\phi}_i^{k+1}) + (1-\alpha^{l,k+1})\hat{\boldsymbol{h}}^{l,k},$$

Weight to measure operation

-> $\{\circ_{\mathrm{op}^{k+1}}\}_{k=0}^{3} = \{\max, \min, \odot, +\}.$



(c) A layer in supernet.

ColdNAS Algorithm

Input: Learning rate β, number of operations to keep K.

1: Construct the supernet by (8) and randomly initialize all parameters $\boldsymbol{\Theta} = \{\{\alpha^{l,k}\}_{k=1,l=0}^{4,L-1}, \boldsymbol{\theta}_E, \boldsymbol{\theta}_A, \boldsymbol{\theta}_P\}$.
2: **while** Not converge **do**
3: **for** Every $T_i \in \mathcal{T}^{\text{train}}$ **do**
4: Calculate $\boldsymbol{\Phi}_i$ by (1). ->調節
5: Calculate $\hat{y}_{i,j}$ for every v_j in $\mathcal{Q}_i$ by (2).
6: Calculate loss $\mathcal{L}_i$ by (10).
7: **end for**
8: $\mathcal{L}^{\text{train}} = \frac{1}{|\mathcal{T}^{\text{train}}|} \sum_{i=1}^{|\mathcal{T}^{\text{train}}|} \mathcal{L}_i$ ->Loss平均
9: Update all parameters $\boldsymbol{\Theta} \leftarrow \boldsymbol{\Theta} - \beta \nabla_{\boldsymbol{\Theta}} \mathcal{L}^{\text{train}}$.
10: **end while**
11: Determine the modulation structure by keeping operations corresponding to Top-K $\alpha^{l,k}$ and remove the others.
12: Construct the model with determined modulation structure and randomly initialize all parameters $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_E, \boldsymbol{\theta}_A, \boldsymbol{\theta}_P\}$.
13: Train the model in the same way as Step 2 ~ 10.
14: **Return:** The trained model.

Outline

- Introduction
- Method
- Experience
- Conclusion

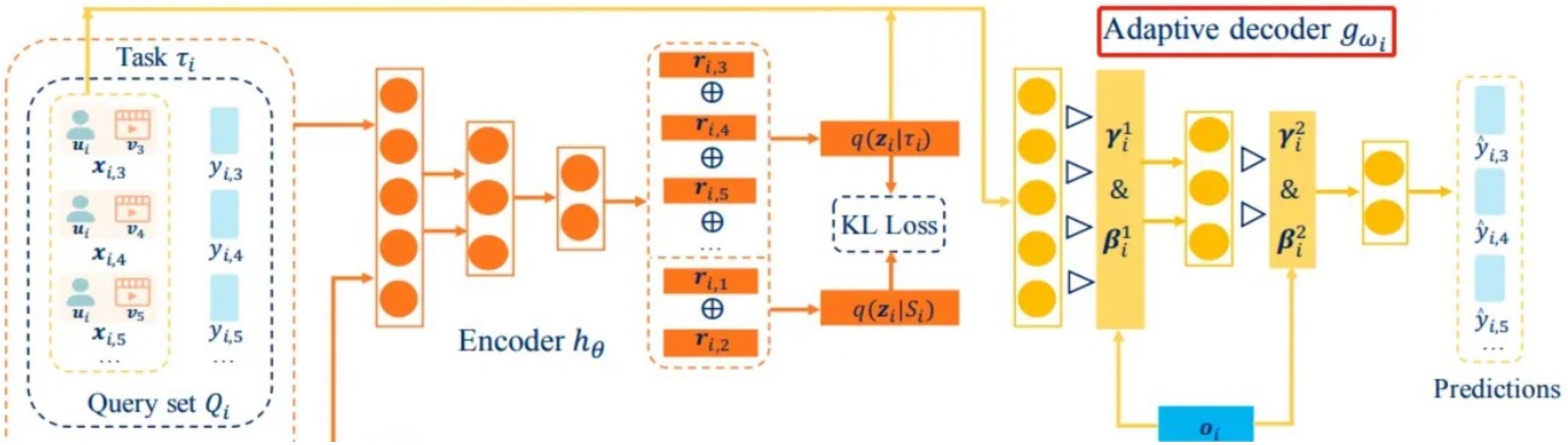
Dataset

Dataset	# User (Cold)	# Item	# Rating	# User Feat.	# Item Feat.
MovieLens	6040 (52.3%)	3706	1000209	4	5
BookCrossing	278858 (18.6%)	271379	1149780	2	3
Last.fm	1872 (15.3%)	3846	42346	1	1

音樂社群

Baseline

- Dropout-Net
- MeLU
- MetaCS
- MetaHIN
- TaNP



Experiment

MSE,MAE是衡量預測值和真實值之間的差值->越低越好
nDCG是衡量推薦列表的相關性->越高越好

Dataset	Metric	DropoutNet	MeLU	MetaCS	MetaHIN	TaNP	ColdNAS-Fixed	ColdNAS
MovieLens	MSE	100.90 (0.70)	95.02 (0.03)	95.05 (0.04)	91.89 (0.06)	89.11 (0.18)	91.05 (0.13)	**87.96** (0.12)
	MAE	85.71 (0.48)	77.38 (0.25)	77.42 (0.26)	75.79 (0.27)	74.78 (0.14)	75.65 (0.30)	**74.29** (0.20)
	nDCG$_3$	69.21 (0.76)	74.43 (0.59)	74.46 (0.78)	74.69 (0.32)	75.60 (0.07)	75.11 (0.09)	**76.16** (0.03)
	nDCG$_5$	68.43 (0.48)	73.52 (0.41)	73.45 (0.56)	73.63 (0.22)	74.29 (0.12)	73.89 (0.12)	**74.74** (0.09)
BookCrossing	MSE	15.38 (0.23)	15.15 (0.02)	15.20 (0.08)	14.76 (0.07)	14.75 (0.05)	14.44 (0.16)	**14.15** (0.08)
	MAE	3.75 (0.01)	3.68 (0.01)	3.66 (0.01)	3.50 (0.01)	3.48 (0.01)	3.49 (0.02)	**3.40** (0.01)
	nDCG$_3$	77.66 (0.18)	77.69 (0.15)	77.68 (0.12)	77.66 (0.19)	77.48 (0.06)	77.65 (0.09)	**77.83** (0.01)
	nDCG$_5$	80.87 (0.15)	81.10 (0.15)	80.97 (0.09)	80.95 (0.04)	81.16 (0.21)	81.12 (0.06)	**81.32** (0.10)
Last.fm	MSE	21.91 (0.38)	21.69 (0.34)	21.68 (0.12)	21.43 (0.23)	21.58 (0.20)	21.62 (0.16)	**20.91** (0.05)
	MAE	43.02 (0.52)	42.28 (1.21)	42.28 (0.76)	42.07 (0.49)	42.15 (0.56)	42.32 (0.34)	**41.78** (0.24)
	nDCG$_3$	75.13 (0.48)	80.15 (2.09)	80.81 (0.97)	82.01 (0.56)	81.03 (0.36)	80.77 (0.32)	**82.80** (0.69)
	nDCG$_5$	69.03 (0.31)	75.03 (0.68)	75.01 (0.64)	75.98 (0.33)	75.98 (0.41)	75.48 (0.21)	**76.77** (0.10)

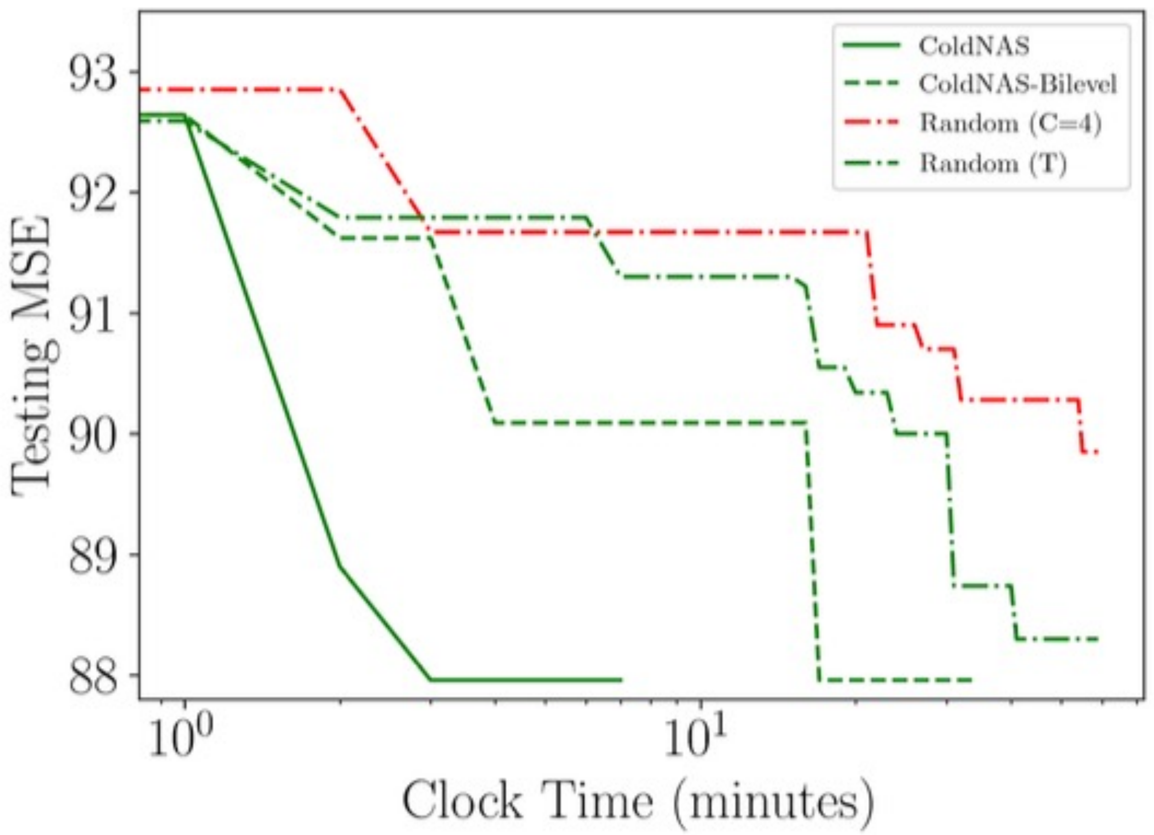
Modulation structure with Top4

	M^0	M^1	M^2	M^3
MovieLens	$\min(\max(\boldsymbol{h}^0, \boldsymbol{\phi}_i^{0,1}), \boldsymbol{\phi}_i^{0,2}) + \boldsymbol{\phi}_i^{0,3}$	$\boldsymbol{h}^1 + \boldsymbol{\phi}_i^{1,1}$	$\boldsymbol{h}^2$	$\boldsymbol{h}^3$
BookCrossing	$\min(\boldsymbol{h}^0, \boldsymbol{\phi}_i^{0,1})$	$\boldsymbol{h}^1 + \boldsymbol{\phi}_i^{1,1}$	$\boldsymbol{h}^2 \odot \boldsymbol{\phi}_i^{2,1} + \boldsymbol{\phi}_i^{2,2}$	$\boldsymbol{h}^3$
Last.fm	$\boldsymbol{h}^0 + \boldsymbol{\phi}_i^{0,1}$	$\boldsymbol{h}^1 + \boldsymbol{\phi}_i^{1,1}$	$\max(\boldsymbol{h}^2, \boldsymbol{\phi}_i^{2,1}) + \boldsymbol{\phi}_i^{2,2}$	$\boldsymbol{h}^3$
ColdNAS-Fixed	$\boldsymbol{h}^0 \odot \boldsymbol{\phi}_i^{0,1} + \boldsymbol{\phi}_i^{0,2}$	$\boldsymbol{h}^1 \odot \boldsymbol{\phi}_i^{1,1} + \boldsymbol{\phi}_i^{1,2}$	$\boldsymbol{h}^2 \odot \boldsymbol{\phi}_i^{2,1} + \boldsymbol{\phi}_i^{2,2}$	$\boldsymbol{h}^3 \odot \boldsymbol{\phi}_i^{3,1} + \boldsymbol{\phi}_i^{3,2}$

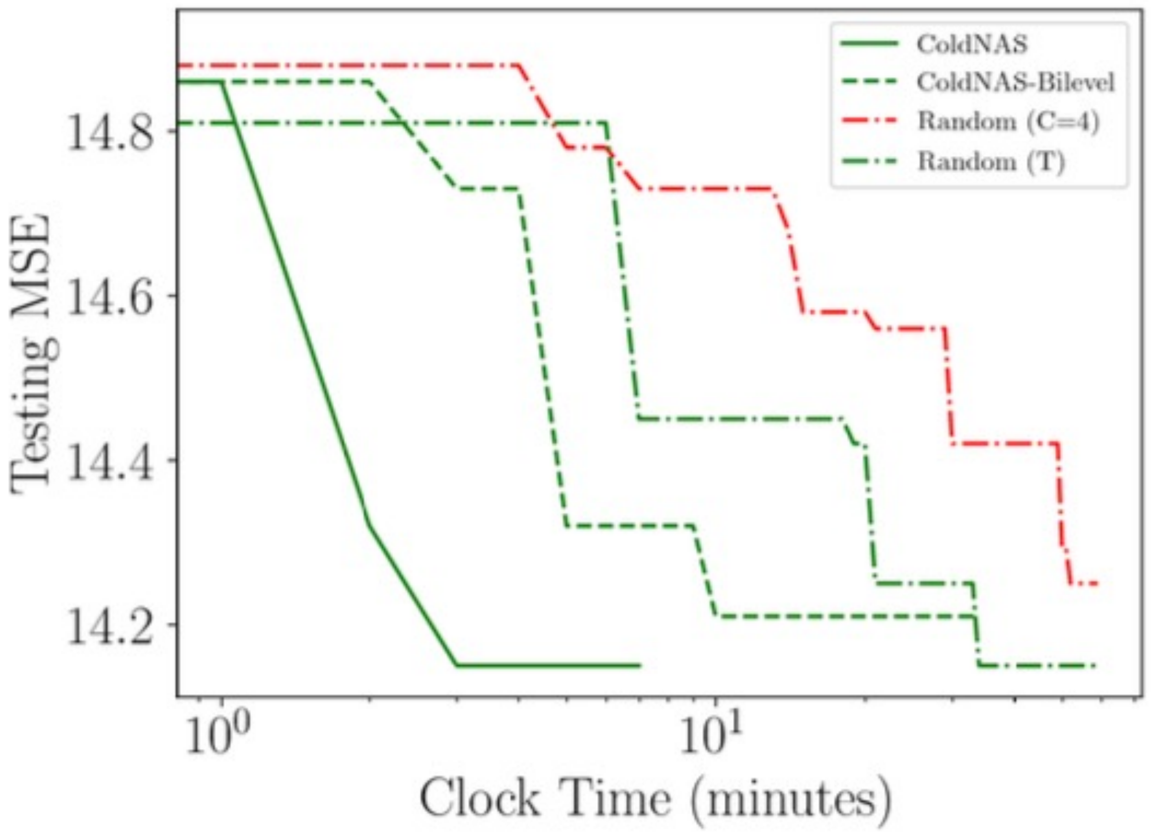
Clock time

Clock time (min)		MovieLens	BookCrossing	Last.fm
TaNP		15.5	44.2	4.1
ColdNAS	Search	16.2	45.5	3.9
	Retrain	12.7	35.5	3.5

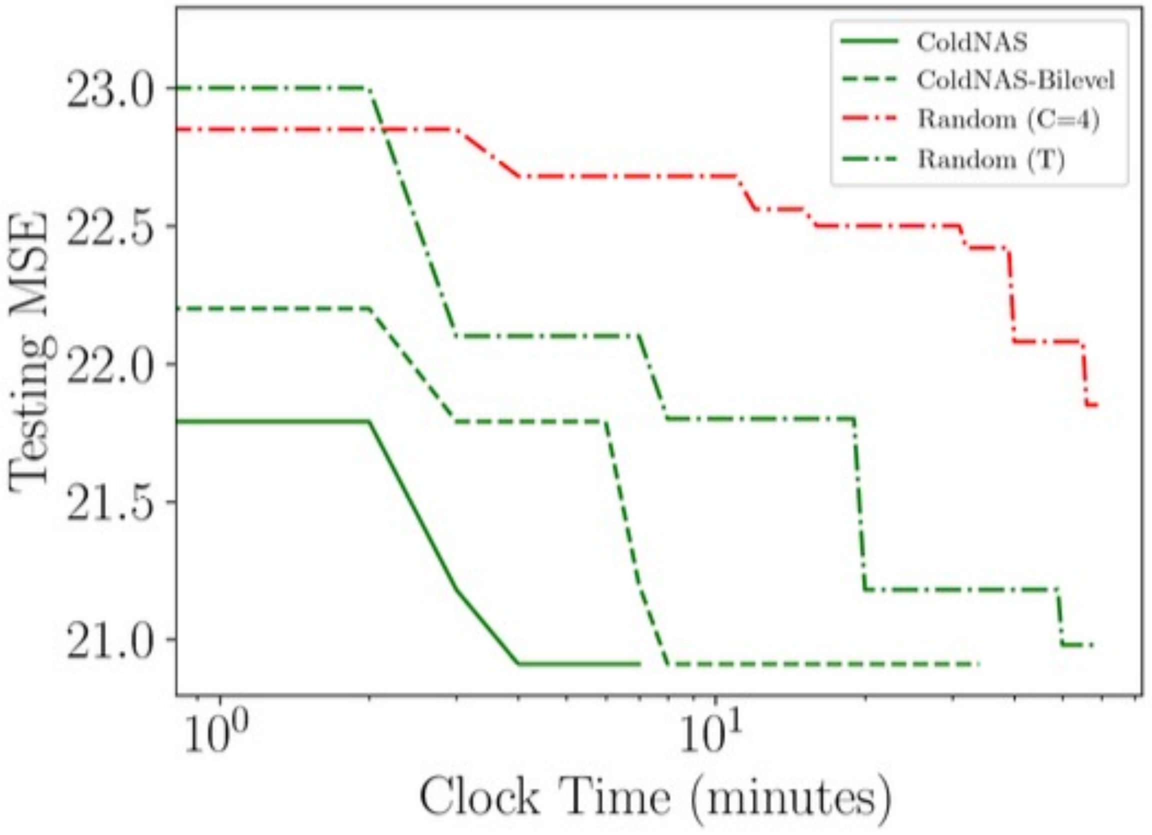
Testing MSE vs clock time



(a) MovieLens.



(b) BookCrossing.



(c) Last.fm.

Outline

- Introduction
- Method
- Experience
- Conclusion

Conclusion

- Propose a modulation framework called ColdNAS for user cold- start recommendation.
- ColdNAS can efficiently find proper modulation structure for different data, which make it easy to be deployed in recommendation systems.