Curriculum Meta-Learning for Next POI Recommendation

Source: KDD '21

Speaker: Po-Jen, Wen

Advisor: Jia-Ling, Koh

Date: 2021/9/13

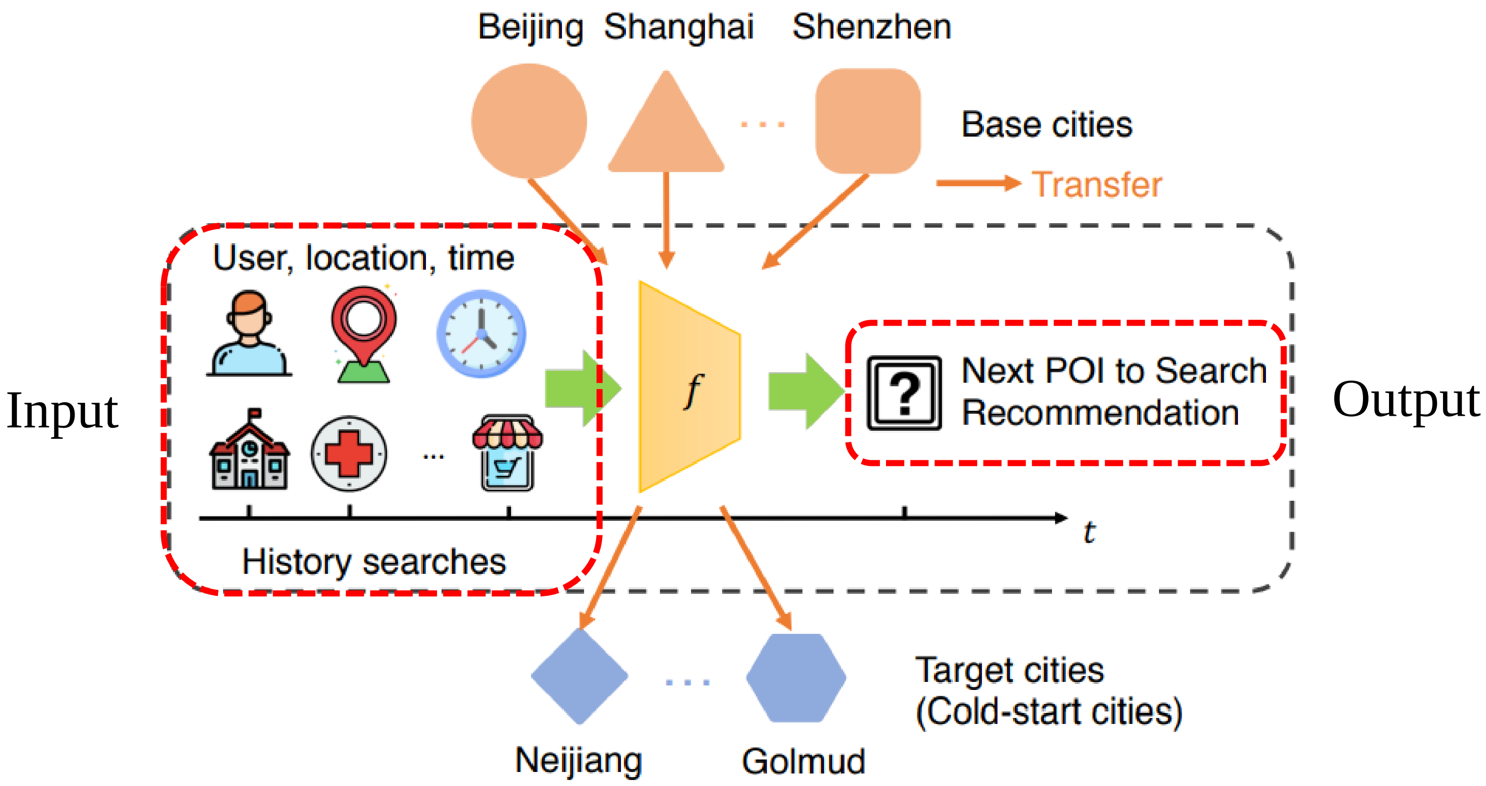
OUTLINE

- Introduction
- Method
- Experiment
- Conclusion

Introduction

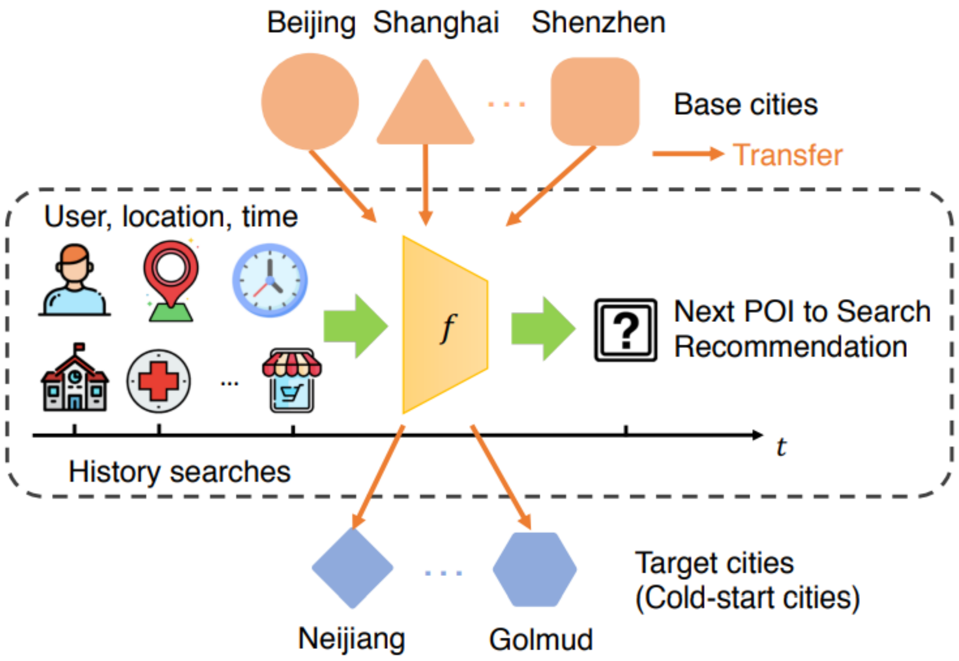
Goal：City-transfer next point-of-interest (POI) to search recommendation

City-transfer：transfer the knowledge learned from base cities (rich data) to target cities (limited data)



Challenges

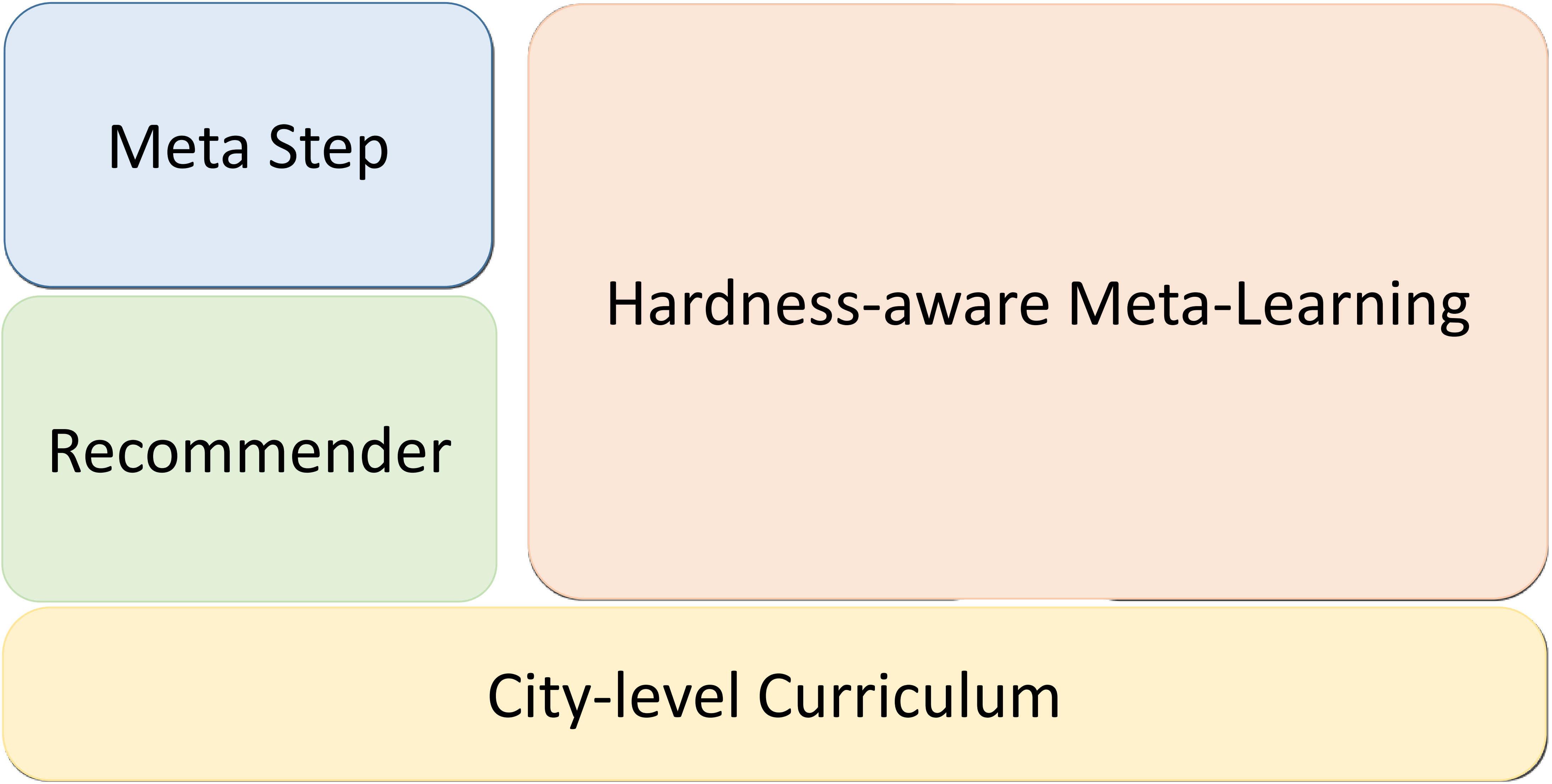
1. The shared data among different cities is extremely limited
    - POIs of different cities have no intersection
    - Users are mostly local residents
2. The map search patterns of various users in different cities are diverse
    - Some types of patterns are trivial to capture
    - Further improvement relies on capturing diverse search patterns among cities and users



OUTLINE

- Introduction
- Method
- Experiment
- Conclusion

Overall framework of CHAML



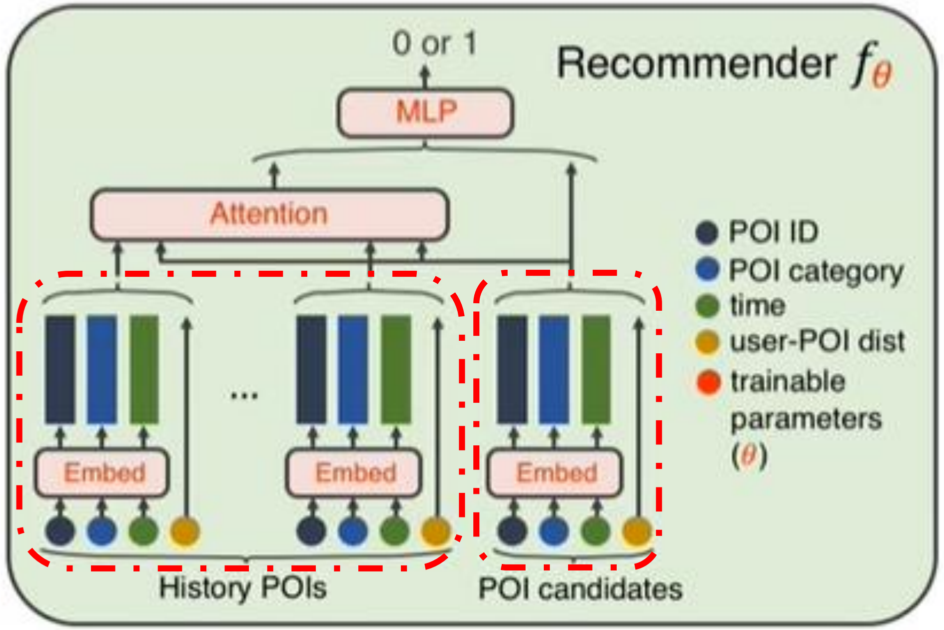
CHAML: Recommender

1. Embedding module $\left(x_i \mapsto (e_{hist}, e_{candi})\right)$

   POI ID v_i, category, hourly timestamp t_i embedding to $E_{id}, E_{cate}, E_{time} \in R^d$

   $E_{id}, E_{cate}, E_{time}$ concatenated to embedding vector e_i

   $e_i \leftarrow [e_i; \frac{d_i - d_{mean}}{d_{std}}]$ d_i user與POI之間的距離, d_{mean}城市平均距離, d_{std}距離標準差



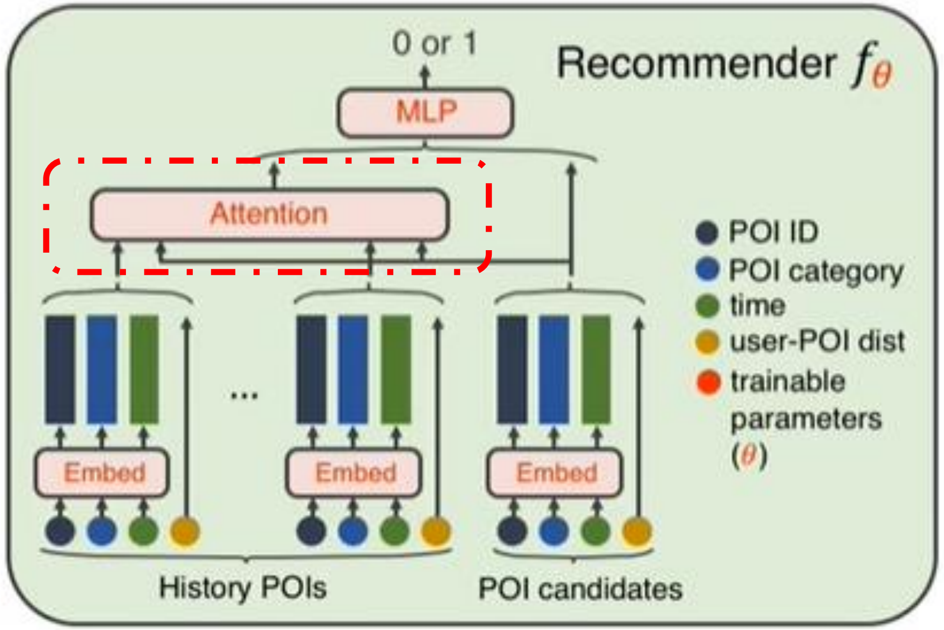
e_{hist} e_{candi}

CHAML: Recommender

2. Attention module ($(e_{hist}, e_{candi}) \mapsto h$)

$$h = attention(e_{candi}, e_{hist})$$

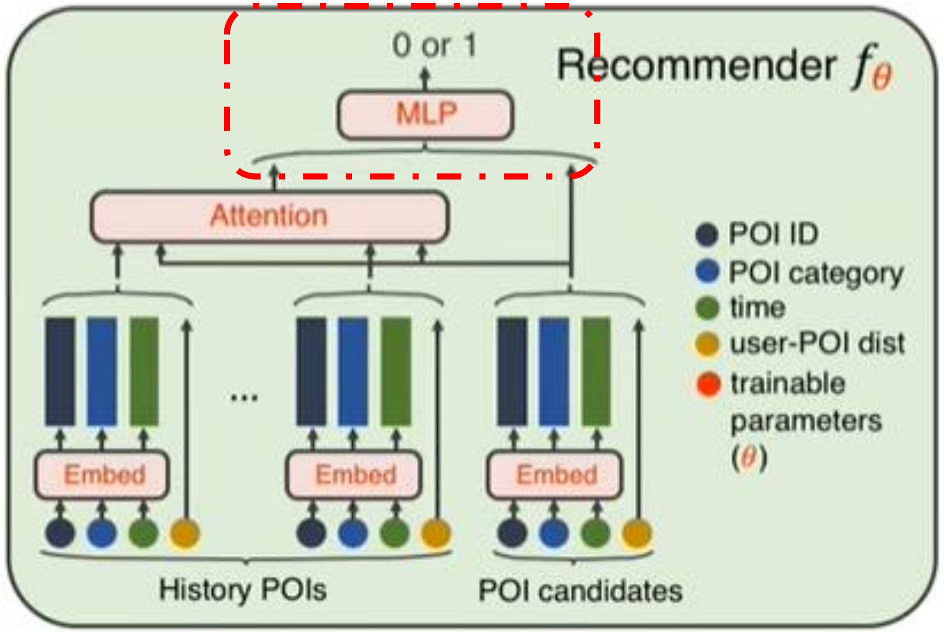
$$attention(K, V) = softmax\big(MLP_{att}([K; V; K - V; K \cdot V])\big)V$$



CHAML: Recommender

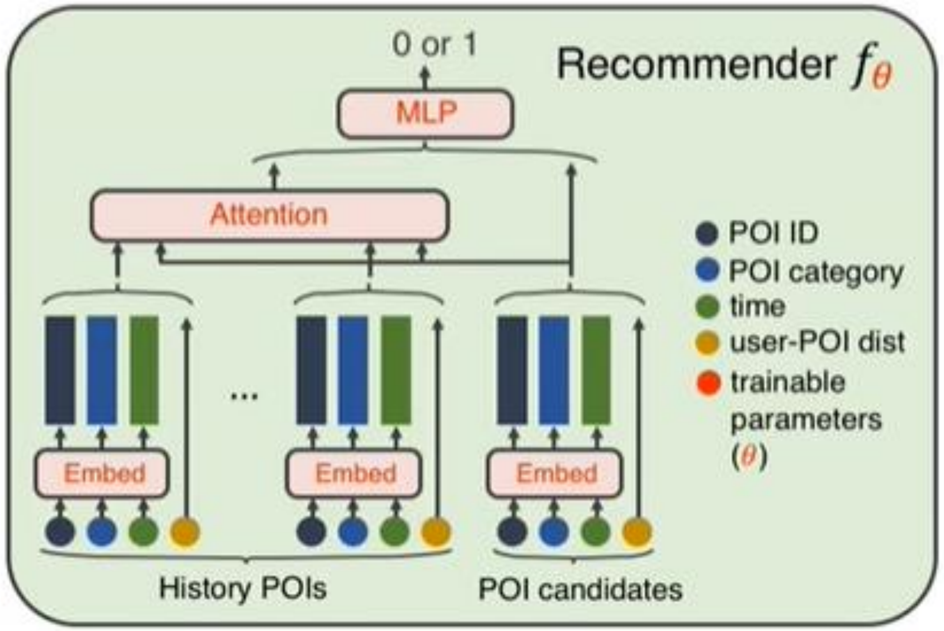
3. Output module ($(h, e_{candi}) \mapsto \hat{y}_i$)

$$\hat{y}_i = MLP_{rec}([h; e_{candi}])$$



CHAML: Recommender

Meta-Learner: to learn the initialization of θ to aid diverse target cities and users



CHAML: Meta Step

Support set D^{spt}: chronologically select the first k samples of each user into D^{spt}

Query set D^{qry}: rest into D^{qry}

Samples:

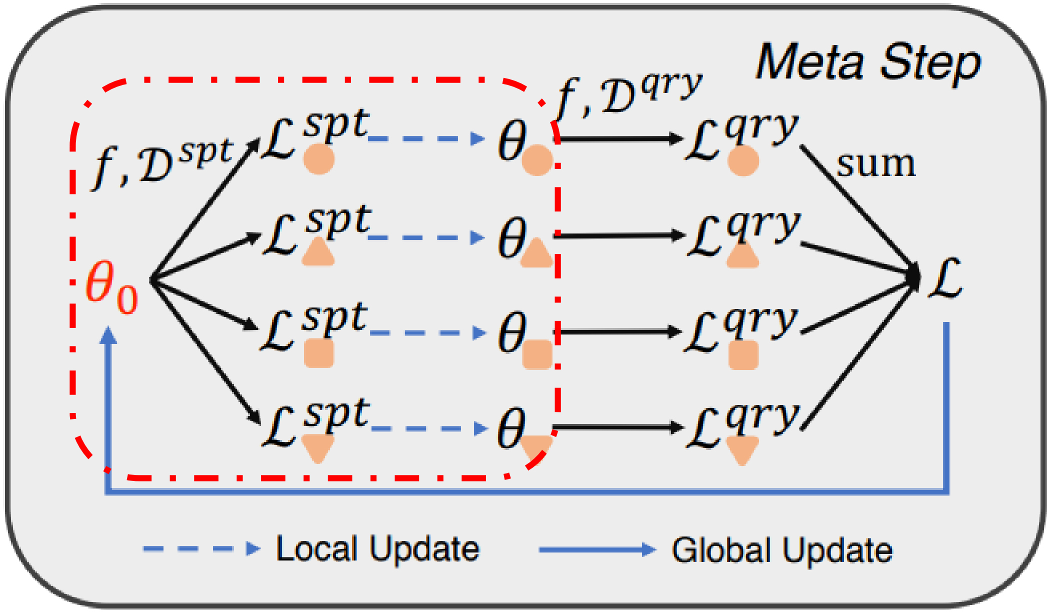
Positive sample:

$$x_i = \left(u, hist_u^i, r_i\right), y_i = 1$$

$$hist_u^i = (r_1, \dots, r_{i-1}), i = m+1, \dots, n$$

Negative sample:

在城市c的POI集合中隨機抓取，且沒有出現在$hist_u^i$中



CHAML: Meta Step

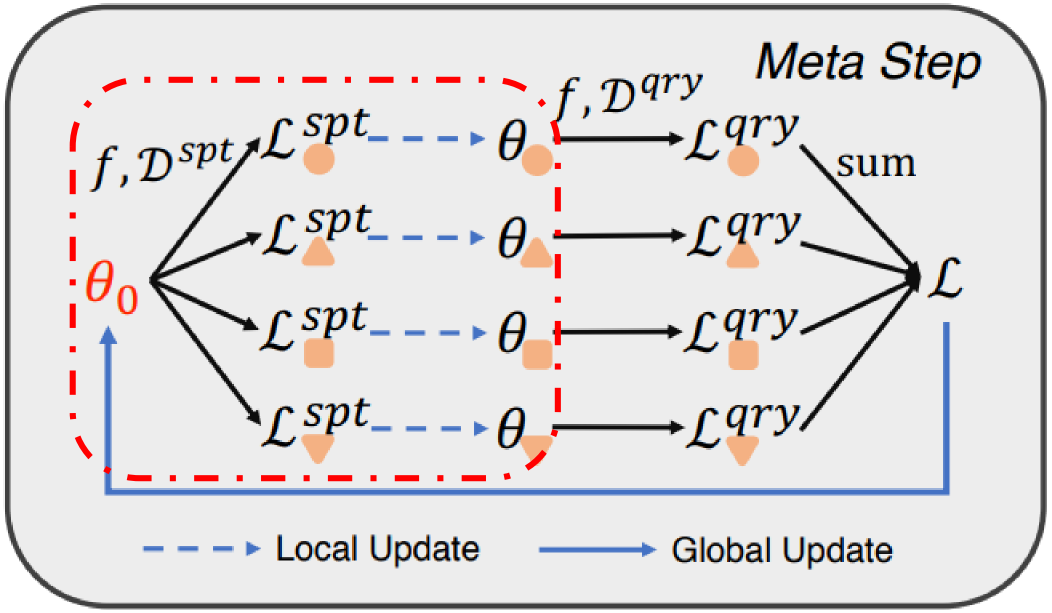
Local update:

$$\theta_c' = \theta - \alpha \nabla_\theta \mathcal{L}_c(f_\theta, \mathcal{D}_c^{spt})$$

α → local learning rate

∇_θ →梯度下降

L_c → cross-entropy loss of binary classification

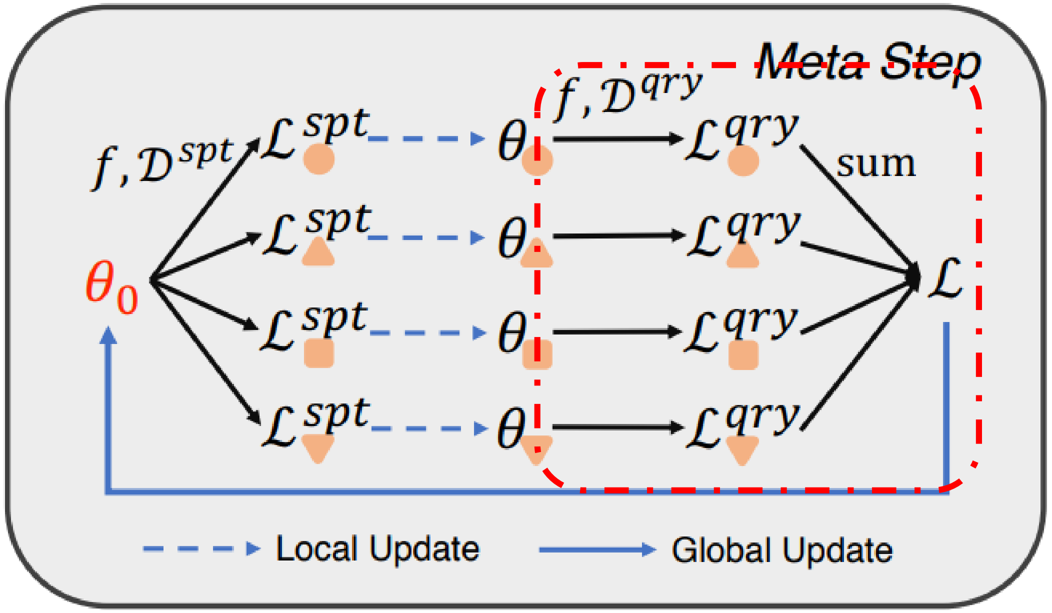


CHAML: Meta Step

Global update:

$$\theta = \theta - \beta \nabla_\theta \sum_{c \in \mathcal{B}} \mathcal{L}_c(f_{\theta'_c}, \mathcal{D}_c^{qry})$$

$\beta \rightarrow$ global learning rate

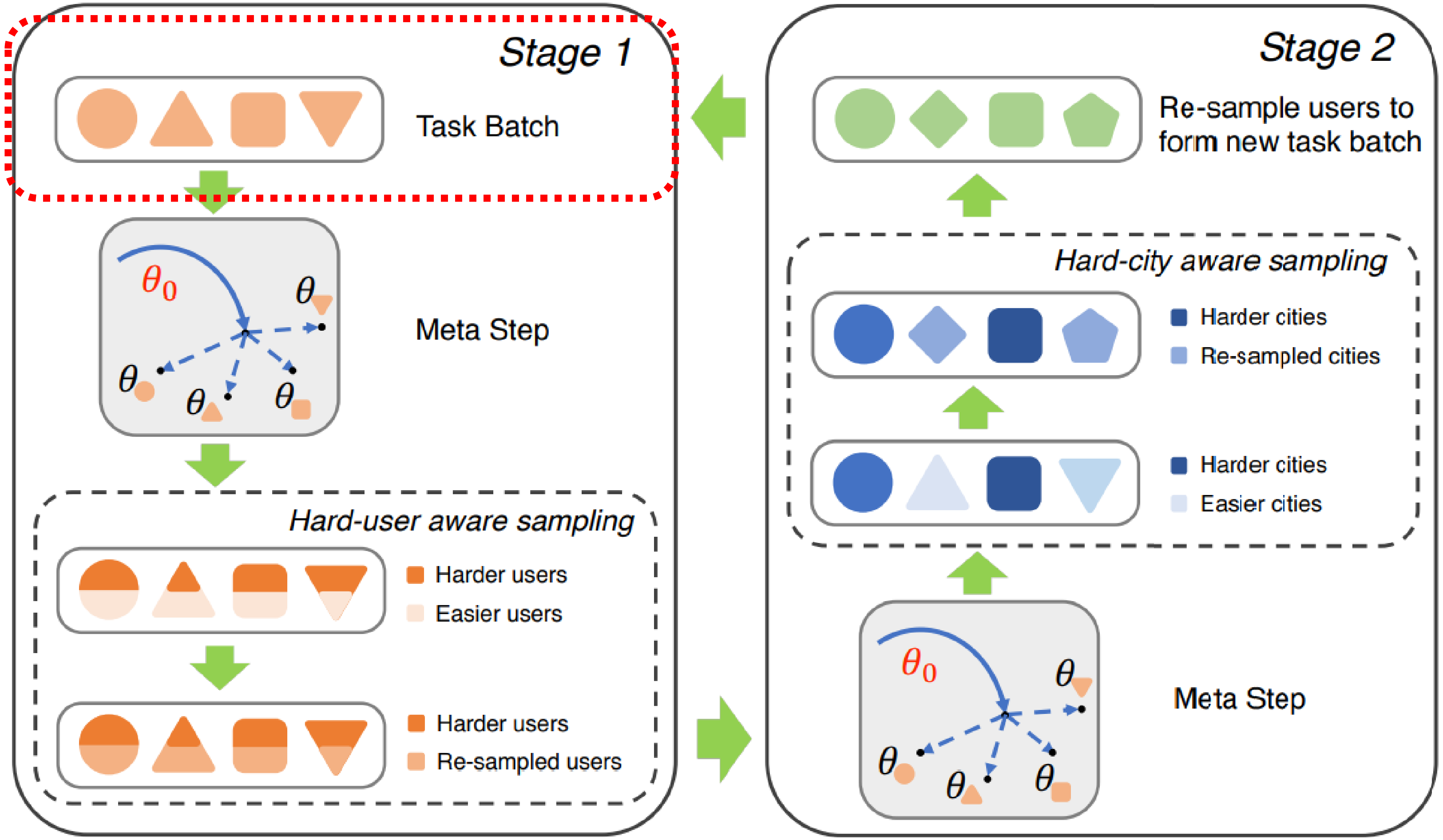


CHAML: Hardness-aware meta-learning

執行M次 (Max step of iterations)

T_{hard_city}

T_{hard_user}

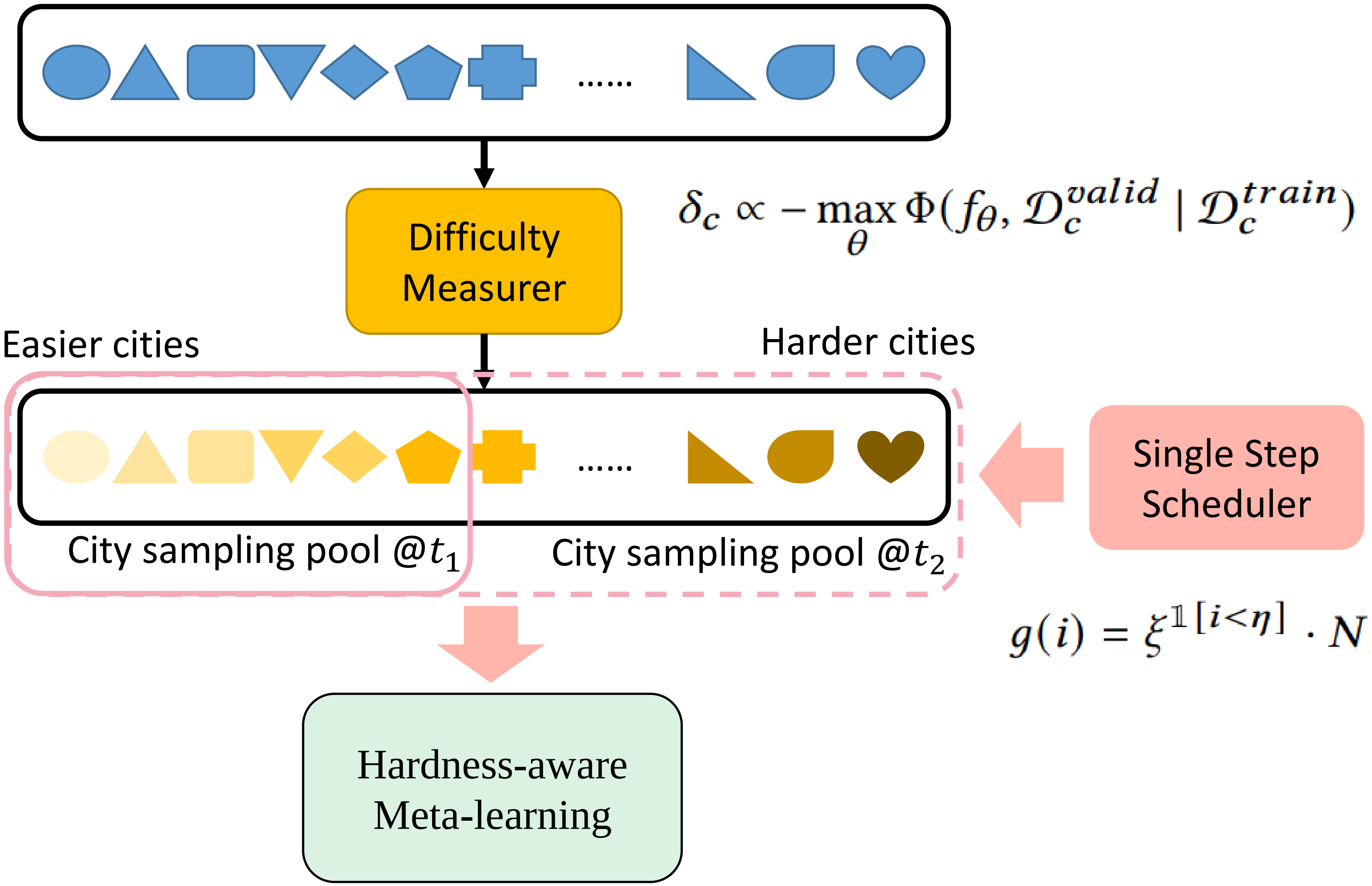


T_{hard_city}

T_{hard_user}

Keep $k_u B_u$ hardest users (lowest accuracies)
re-sample new user

CHAML: City-level Curriculum



$$\delta_c \propto -\max_{\theta} \Phi(f_\theta, \mathcal{D}_c^{valid} \mid \mathcal{D}_c^{train})$$

$\Phi \rightarrow best\ valid\ score$

$$g(i) = \xi^{\mathbb{1}[i<\eta]} \cdot N$$

$\xi \rightarrow$ the fraction of cities

$\eta \rightarrow$ the number of iterations

$1 \rightarrow$ indicator function

$$1_A(x) := \begin{cases} 1, x \in A \\ 0, x \notin A \end{cases}$$

OUTLINE

- Introduction
- Method
- Experiment
- Conclusion

Datasets and baselines

- Datasets: two from Baidu Maps
  - Differ in city amount and tiers

Dataset	#Base cities/users	#Target cities/users	#POIs
MapSmall	8 / 160,000	6 / 3,000	477,218
MapLarge	72/1,090,070	43/19,672	2,294,879

- Baselines: recommenders × transfer methods
  - Recommenders:
    popular models and state-of-the-art (SOTA) next POI recommendation models
  - Transfer methods:
    - No transfer (None): only train on the support set of the cold-start cities
    - Pretrain and fine-tune (FT): pretrain on all base cities, fine-tune on each cold start city
    - s^2Meta: SOTA meta-learning for cold-start scenarios recommendation

    ----

    - MAML: Recommender f_θ + Meta step
    - HAML: + Hardness-aware meta-learning
    - CHAML: + City-level curriculum

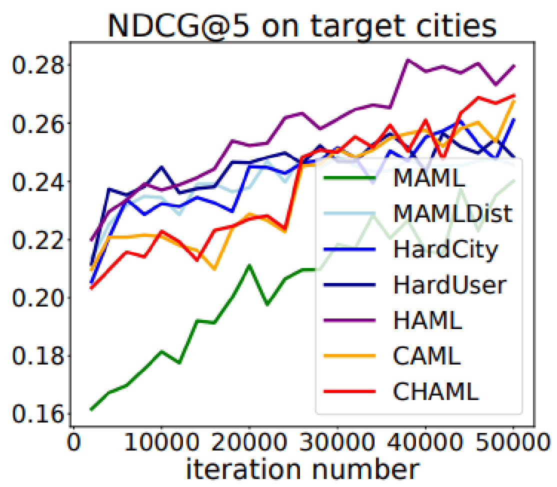
Experiment results

1. Meta-learning is the best transfer method
2. The proposed CHAML framework is the most effective
3. City-level curriculum shows advantage on larger sampling space

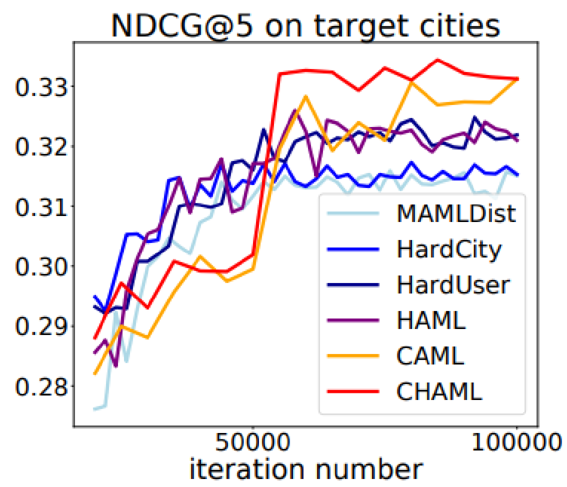
Transfer	Models	*MapSmall*				*MapLarge*			
		H@5	**H@10**	**N@5**	**N@10**	**H@5**	**H@10**	**N@5**	**N@10**
None	NeuMF	0.1002	0.1628	0.0646	0.0847	0.1401	0.2125	0.0966	0.1199
	HGN	0.0630	0.1245	0.0392	0.0589	0.1188	0.1941	0.0791	0.1033
	ATST-LSTM	0.2705	0.3410	0.2177	0.2403	0.2809	0.3616	0.2183	0.2441
	PLSPL	0.1210	0.1855	0.0844	0.1051	0.1824	0.2592	0.1327	0.1573
	iMTL	0.1208	0.2068	0.0779	0.1054	0.1485	0.2323	0.1009	0.1277
	DIN	0.0787	0.1530	0.0466	0.0704	0.0991	0.1649	0.0646	0.0856
FT	NeuMF	0.1402	0.2063	0.0982	0.1192	0.1656	0.2442	0.1166	0.1418
	HGN	0.2867	0.4013	0.1956	0.2325	0.3438	0.4689	0.2459	0.2862
	ATST-LSTM	0.3218	0.4127	0.2497	0.2788	0.3332	0.4391	0.2513	0.2854
	PLSPL	0.2878	0.3952	0.2060	0.2407	0.3700	0.4791	0.2764	0.3117
	iMTL	0.2820	0.3698	0.2256	0.2538	0.4108	0.5215	0.3254	0.3610
	DIN	0.2985	0.4272	0.1989	0.2405	0.3232	0.4560	0.2246	0.2673
Meta	MAML	0.3478	0.4503	0.2402	0.2736	0.4007	0.4790	0.3252	0.3505
	s^2Meta	0.3395	0.4593	0.2322	0.2709	0.4145	0.5233	0.3114	0.3466
	HAML (Ours)	**0.4152**	**0.5632**	**0.2806**	**0.3286**	0.4465	0.5747	0.3238	0.3653
	CHAML (Ours)	0.4008	0.5543	0.2695	0.3191	**0.4571**	**0.5872**	**0.3320**	**0.3742**

Ablation study

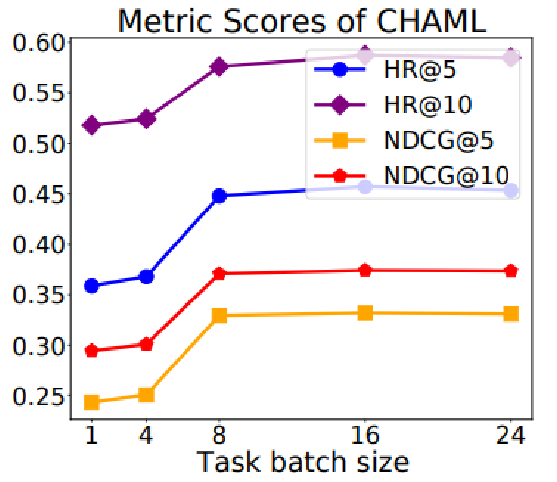
- (a)(b): on two datasets, all the designed components are helpful
- (c): CHAML is sensitive to the batch size of sampled cities
- (d):CHAML is not sensitive to the number of local update steps



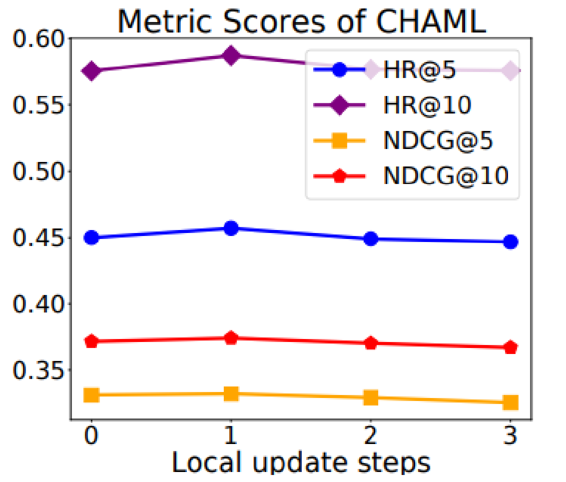
(a) NDCG@5 on *MapSmall*



(b) NDCG@5 on *MapLarge*



(c) Sensitivity on task batch size B_c



(d) Sensitivity on local-update steps

OUTLINE

- Introduction
- Method
- Experiment
- Conclusion

Conclusion

- Propose CHAML, an algorithm of curriculum meta-learning to solve the city-transfer next POI to search recommendation problem
- Effectiveness validated on two large-scale Baidu Maps Datasets