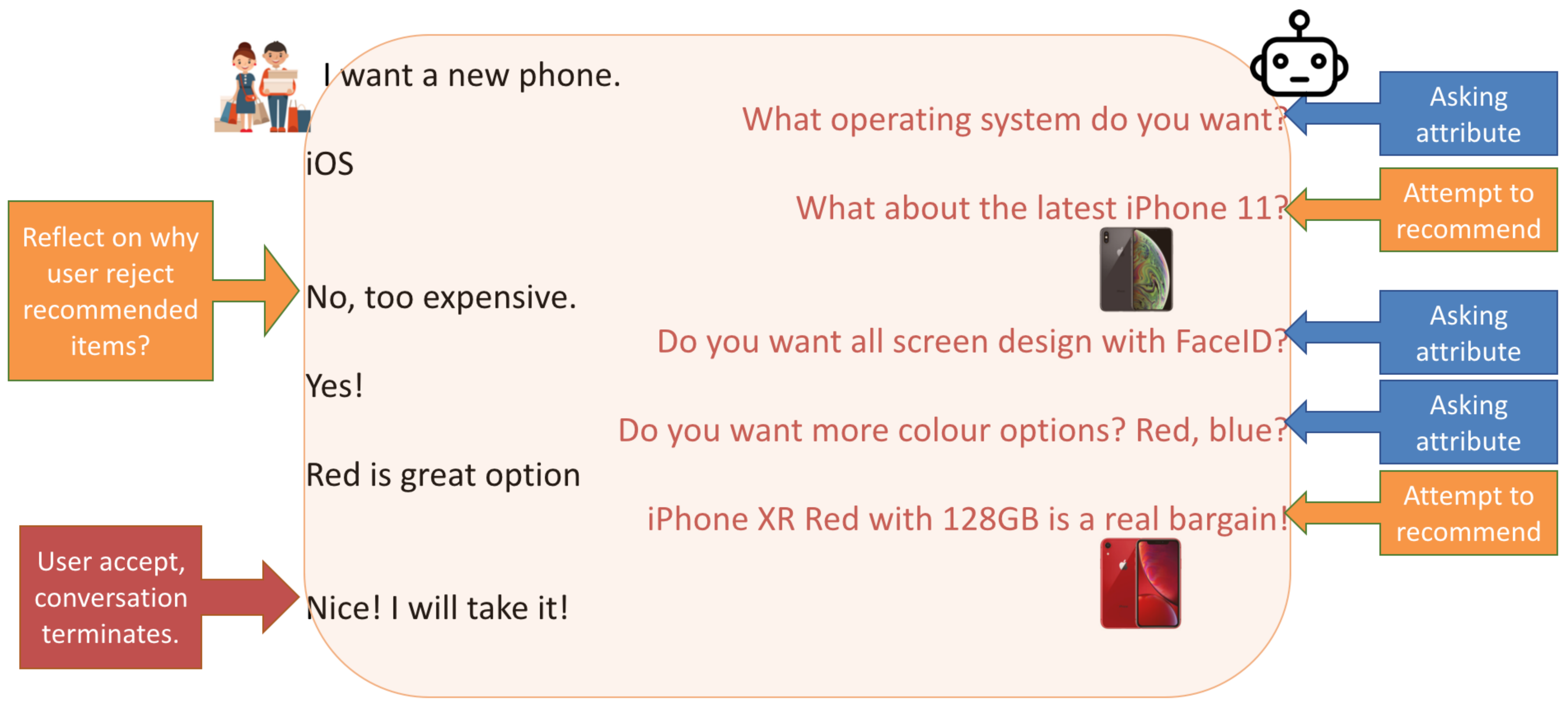
Estimation–Action–Reflection:

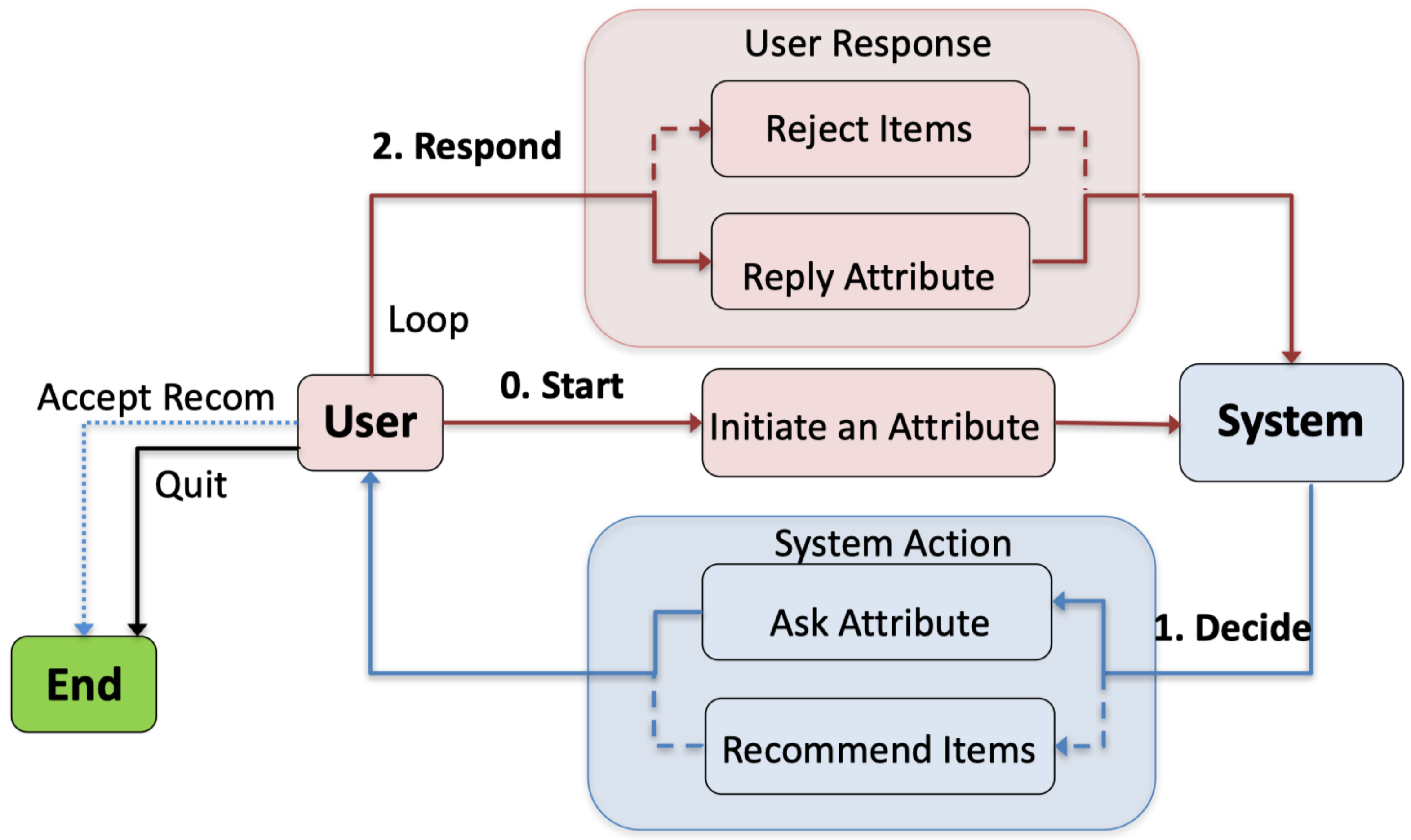
Towards Deep Interaction Between Conversational and Recommender Systems

Wenqiang Lei, Xiangnan He, Yisong Miao, Qingyun Wu, Richang Hong, Min-Yen Kan, Tat-Seng Chua
{wenqianglei, xiangnanhe, miaoyisong}@gmail. com, qw2ky@virginia. edu, hongrc@hfut. edu. cn, {kanmy,chuats}@comp. nus. edu. sg

Introduction to Conversational Recommender System (CRS)



Multi-round CRS scenario



Motivation

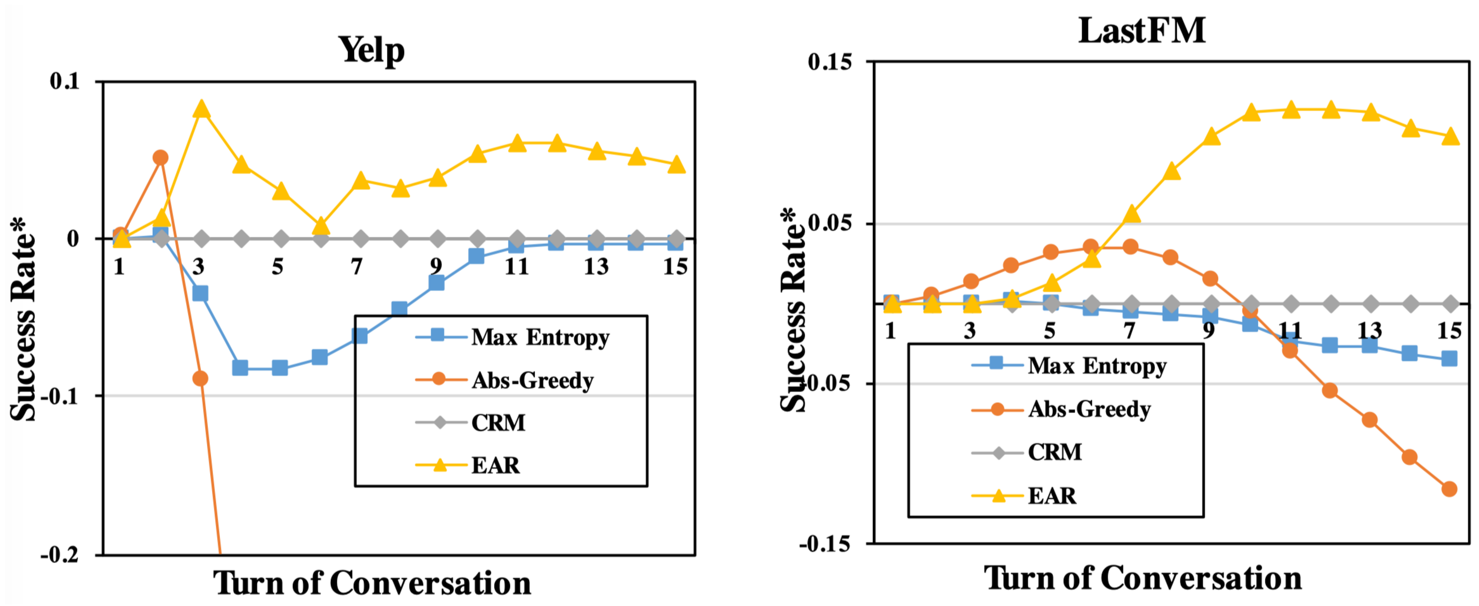
	Ask Attribute	Focus on Strategy	Multi-Round	Unify CC and RC*
Bandit	✗	✗	✓	✗
CRM(SIGIR'18)	✓	✓	✗	✗
Q&R(KDD'18)	✓	✗	✗	✗
EAR	✓	✓	✓	✓

Experiment Setup

	#users	#items	#interaction	#attributes	type of attributes
Yelp	27,675	70,311	1,368,606	590	Enumerated*
LastFM	1,801	7,432	76,693	33	Binary*

* Enumerated Question: Wine: {Red Wine, White Wine, Whiskey}, Binary Question: Classic, Pop, Rock …

Main Experiment Result



Success Rate* indicates the performance difference between models against strongest baseline, CRM

Stage 1
Evaluation metric: AUC score

	LastFM		Yelp	
	Item	Attribute	Item	Attribute
FM	0.521	0.727	0.834	0.654
FM+A	0.724	0.629	0.866	0.638
FM+A+MT	**0.742***	**0.760***	**0.870***	**0.896***

Stage 2 & 3
Evaluation metric: Success Rate @ t
Average Turn of Conversation

	Yelp				LastFM			
	SR@5	SR@10	SR@15	AT	SR@5	SR@10	SR@15	AT
$-\mathbf{s}_{ent}$	0.614	0.895	0.969	4.81	**0.051**	0.190	0.346	12.82
$-\mathbf{s}_{pre}$	0.596	0.857	0.959	5.06	0.024	0.231	0.407	12.55
$-\mathbf{s}_{his}$	0.624	0.894	0.949	4.79	0.021	0.236	0.424	12.50
$-\mathbf{s}_{len}$	0.550	0.846	0.952	5.44	0.013	0.230	0.416	12.56
EAR	**0.629***	**0.907***	**0.971***	**4.71***	0.020	**0.243***	**0.429***	**12.45***

	Yelp				LastFM			
	SR@5	SR@10	SR@15	AT	SR@5	SR@10	SR@15	AT
-update	0.629	0.905	0.970	4.72	0.020	0.217	0.393	12.67
EAR	**0.629**	**0.907**	**0.971**	**4.71**	**0.020**	**0.243***	**0.429***	**12.45***

Conclusion

- We formulate CRS in a multi-round scenario and propose EAR, towards the deep interaction between CC and RC.
- Our FM+A+MT has the best performance on Estimation stage for item prediction and attribute prediction. The Action is bettered by statistics from CC. The reflection stage is especially useful when offline AUC is lower.

Stage 1: Estimation

We collect data from CC to train RC

- Attribute-aware BPR for Item Prediction

$$L_{item} = \sum_{(u,v,v')\in\mathcal{D}_1} -\ln\sigma\left(\hat{y}(u,v,\mathcal{P}_u) - \hat{y}(u,v',\mathcal{P}_u)\right) + \sum_{(u,v,v')\in\mathcal{D}_2} -\ln\sigma\left(\hat{y}(u,v,\mathcal{P}_u) - \hat{y}(u,v',\mathcal{P}_u)\right) + \lambda_\Theta\|\Theta\|^2$$

$$\hat{y}(u,v,\mathcal{P}_u) = \mathbf{u}^T\mathbf{v} + \sum_{p_i\in\mathcal{P}_u} \mathbf{v}^T\mathbf{p}_i$$

Two types of negative samples for BPR:
- D1: Randomly sampled
- D2: Sampled from candidate items
- u: user, v: item, Pu: Known attributes.

- Attribute Preference Prediction

$$L_{attr} = \sum_{(u,p,p')\in\mathcal{D}_3} -\ln\sigma\left(\hat{g}(p|u,\mathcal{P}_u) - \hat{g}(p'|u,\mathcal{P}_u)\right) + \lambda_\Theta\|\Theta\|^2$$

$$\hat{g}(p|u,\mathcal{P}_u) = \mathbf{u}^T\mathbf{p} + \sum_{p_i\in\mathcal{P}_u} \mathbf{p}^T\mathbf{p}_i$$

BPR for paired attributes:
- p: ground truth attributes in this session
- p' : sampled negative attributes

$$L = L_{item} + L_{attr}$$

Jointly Optimise two tasks

Stage 2: Action

We leverage statistics from RC to decide CC's strategy

- Reinforcement learning: Policy Gradient to find best strategy.
- Action space: |P| + 1

State vector components	Meaning	Source
$S_{entropy}$	Encode the entropy of each attribute	RC
$S_{preference}$	Encode estimated preference of each attribute	RC
$S_{history}$	Encode the conversation history	CC
S_{length}	Encode the candidate item list length	RC

Stage 3: Reflection

We leverage information from CC to adjust RC's estimation towards user presence

- Rejected items as negative samples

$$L_{ref} = \sum_{(u,v,v')\in\mathcal{D}_4} -\ln\sigma\left(\hat{y}(u,v,\mathcal{P}_u) - \hat{y}(u,v',\mathcal{P}_u)\right) + \lambda_\Theta\|\Theta\|^2$$

Two types of negative samples for BPR:
- v: original positive sample
- v': recently rejected items