
بسم الله الرحمن الرحيم

دانشگاه صنعتی اصفهان

دانشکده برق و کامپیوتر

محمدعلی میرزایی بادیزی

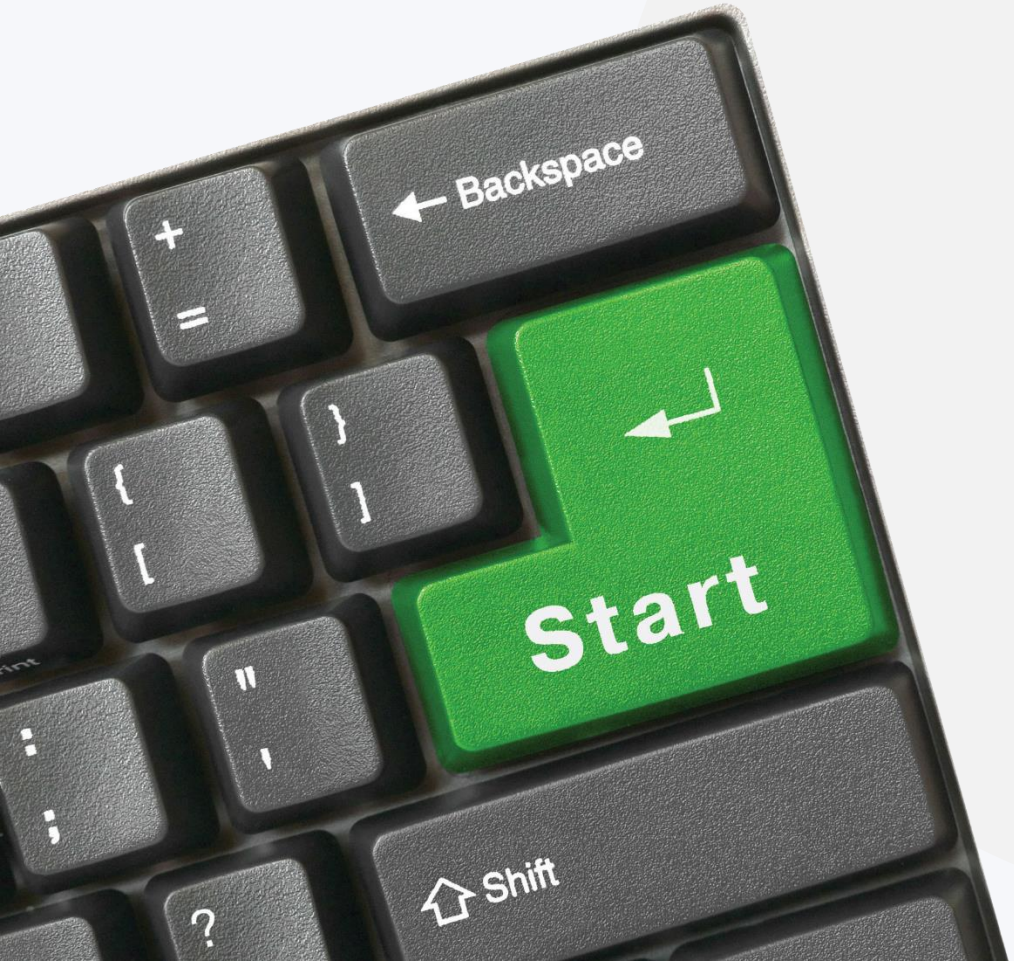
استاد راهنما: دکتر پالهنگ

تسریع یادگیری مشارکتی در سیستم‌های چند عاملی

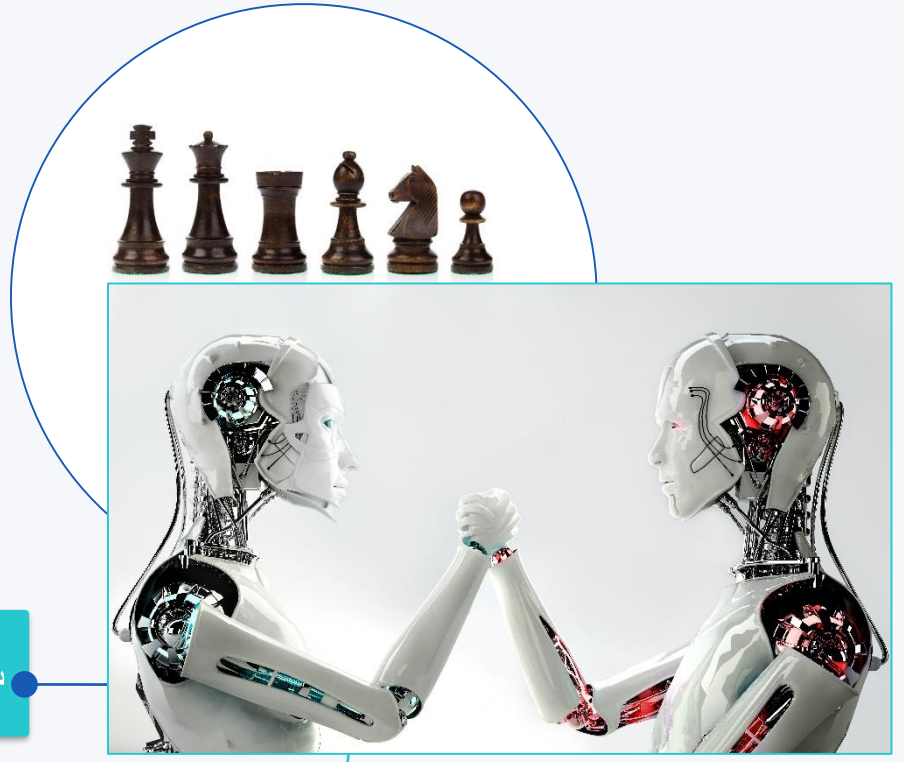
با بهره گیری از کوتاهترین مسیر تجربه شده

فهرست مطالب

مقدمه

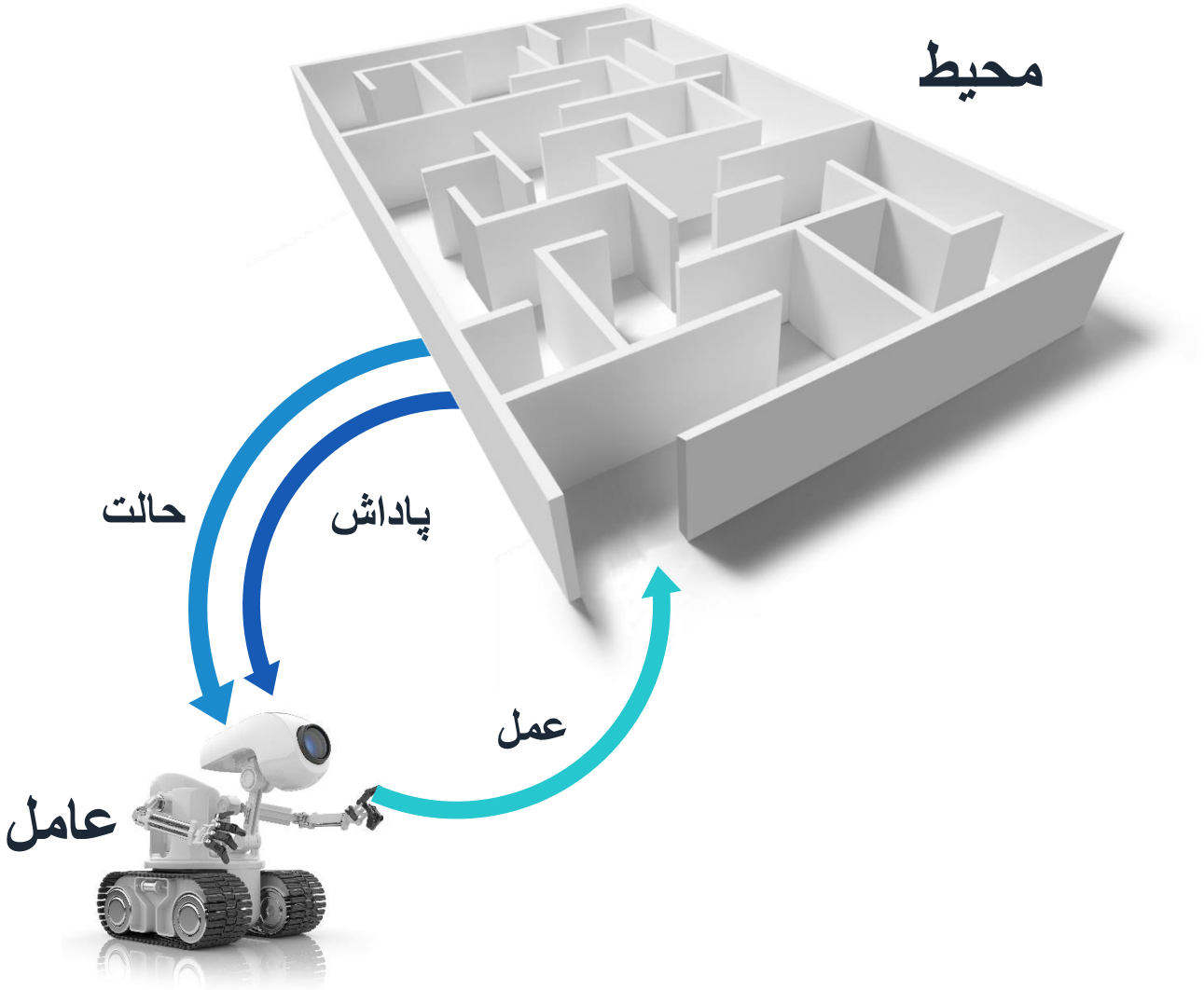


جایگاه پژوهش انجام شده



- هوش مصنوعی
- سیستم‌های توزیع‌شده
  - هوش مصنوعی توزیع شده
    - حل مسئله توزیع شده
    - سیستم های چند عاملی
      - سیستم های چند عامله رقابتی
      - سیستم های چند عامله مشارکتی
        - یادگیری همزمان
        - یادگیری تیمی

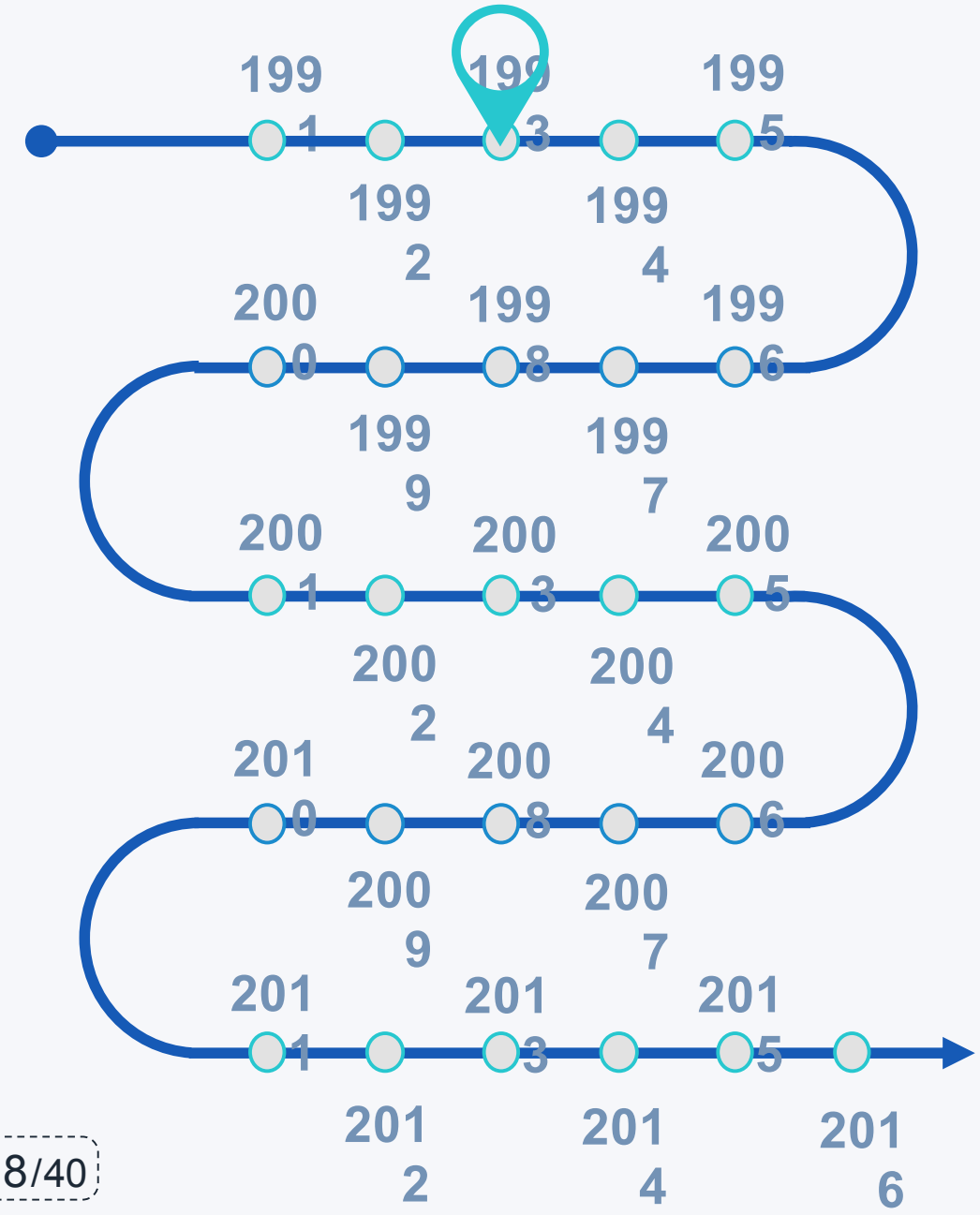
یادگیری تقویتی



Q learning

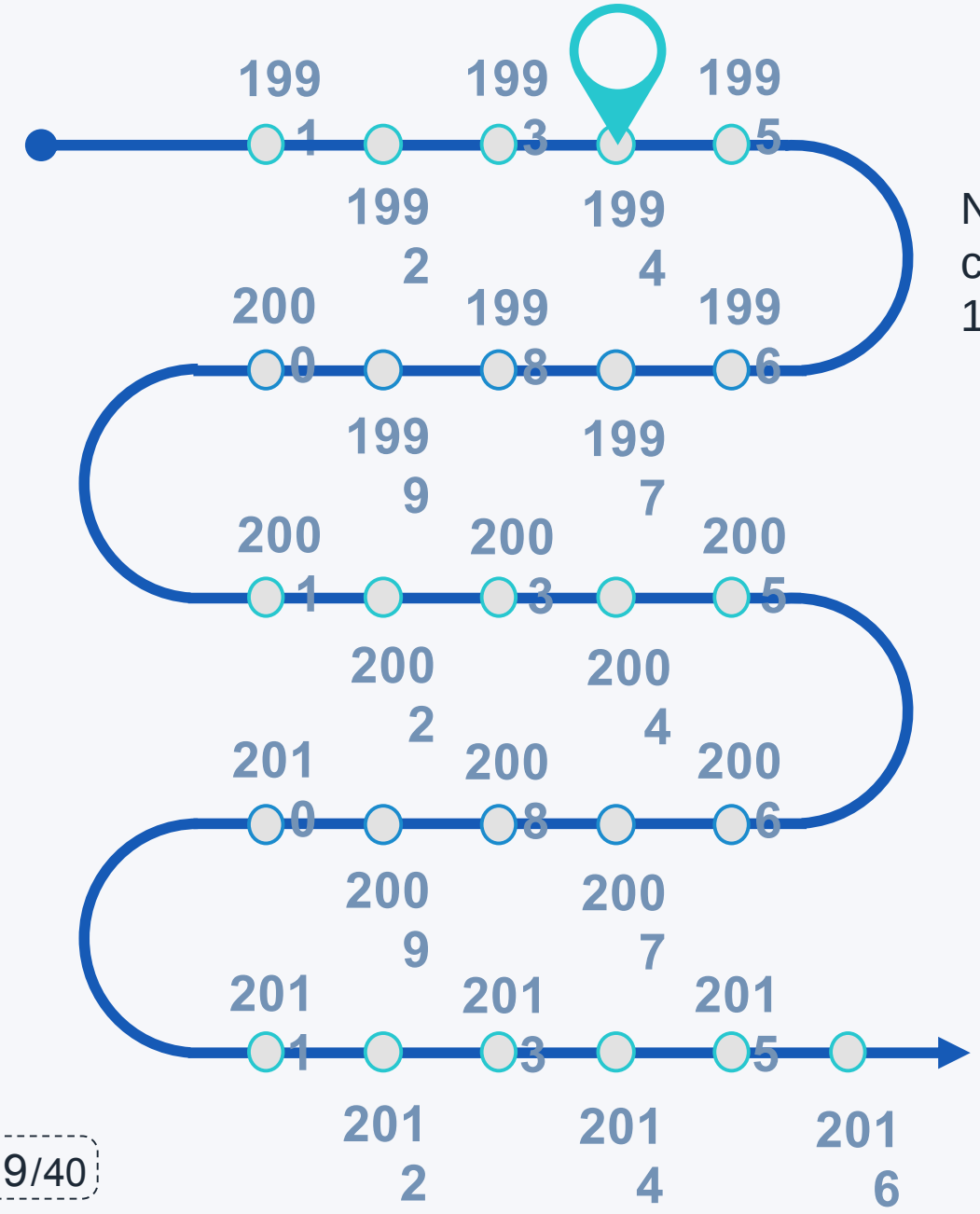
1				
2				
3				
4				
5				
6				

ارائه روشهای پیشین



M. Tan, “Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents,” Proc. Tenth Int. Conf. Mach. Learn., pp. 330–337, 1993.

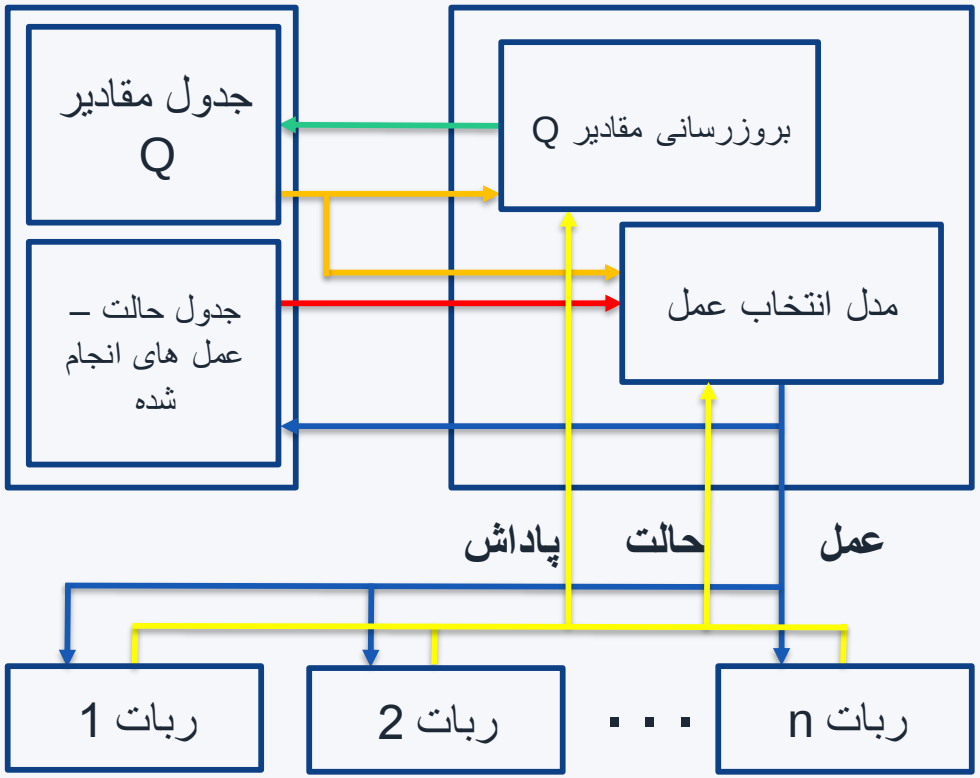
اگر مشارکت به‌خوبی پیاده‌سازی شود هر عامل می‌تواند از تجربیات عامل‌های دیگر استفاده بهینه نماید .

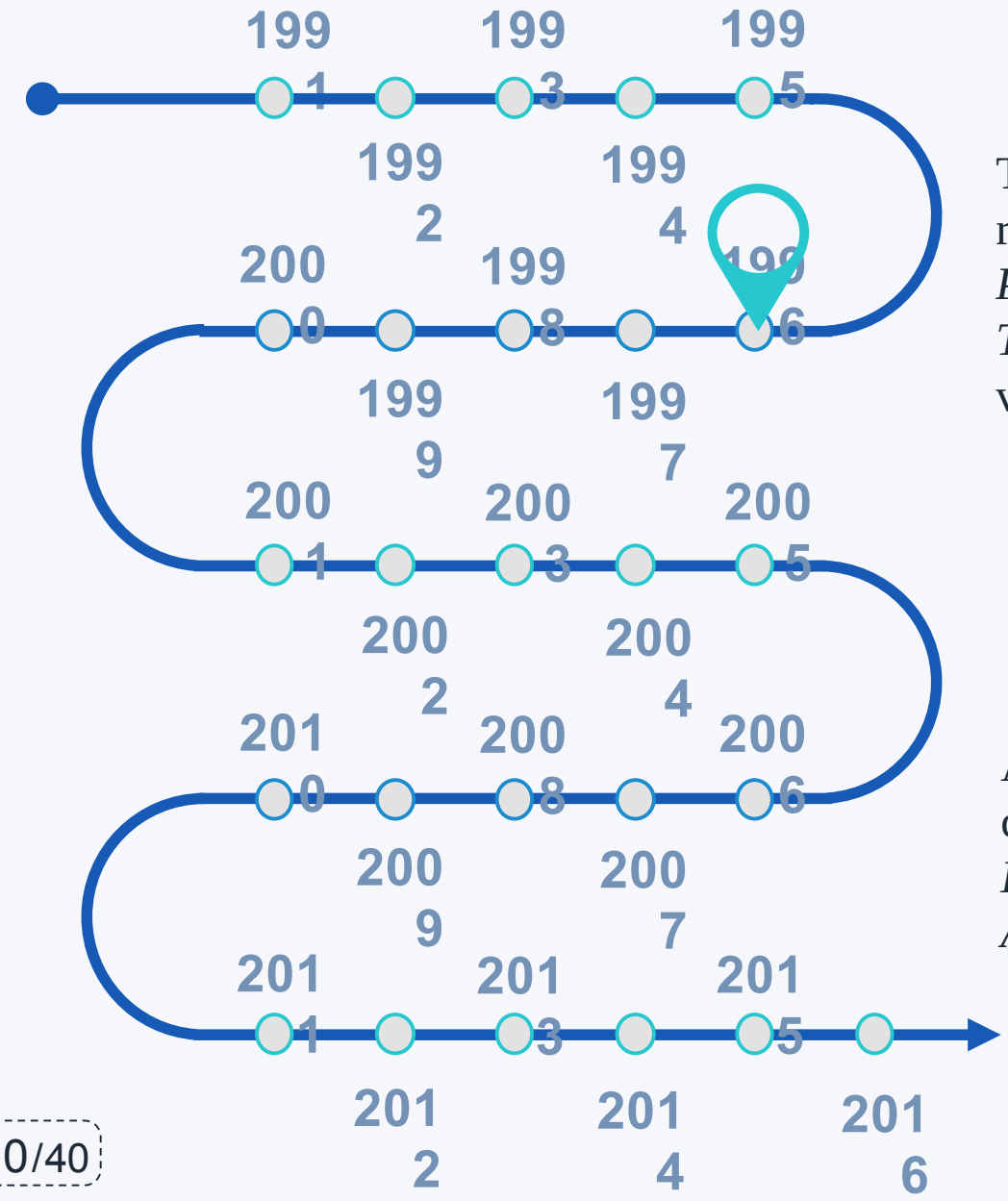


تخته سیاه

Blackboard

N. Carver and V. Lesser, “Evolution of blackboard control architectures,” *Expert Syst. Appl.*, vol. 7, no. 1, pp. 1–30, Jan. 1994.





تقلید

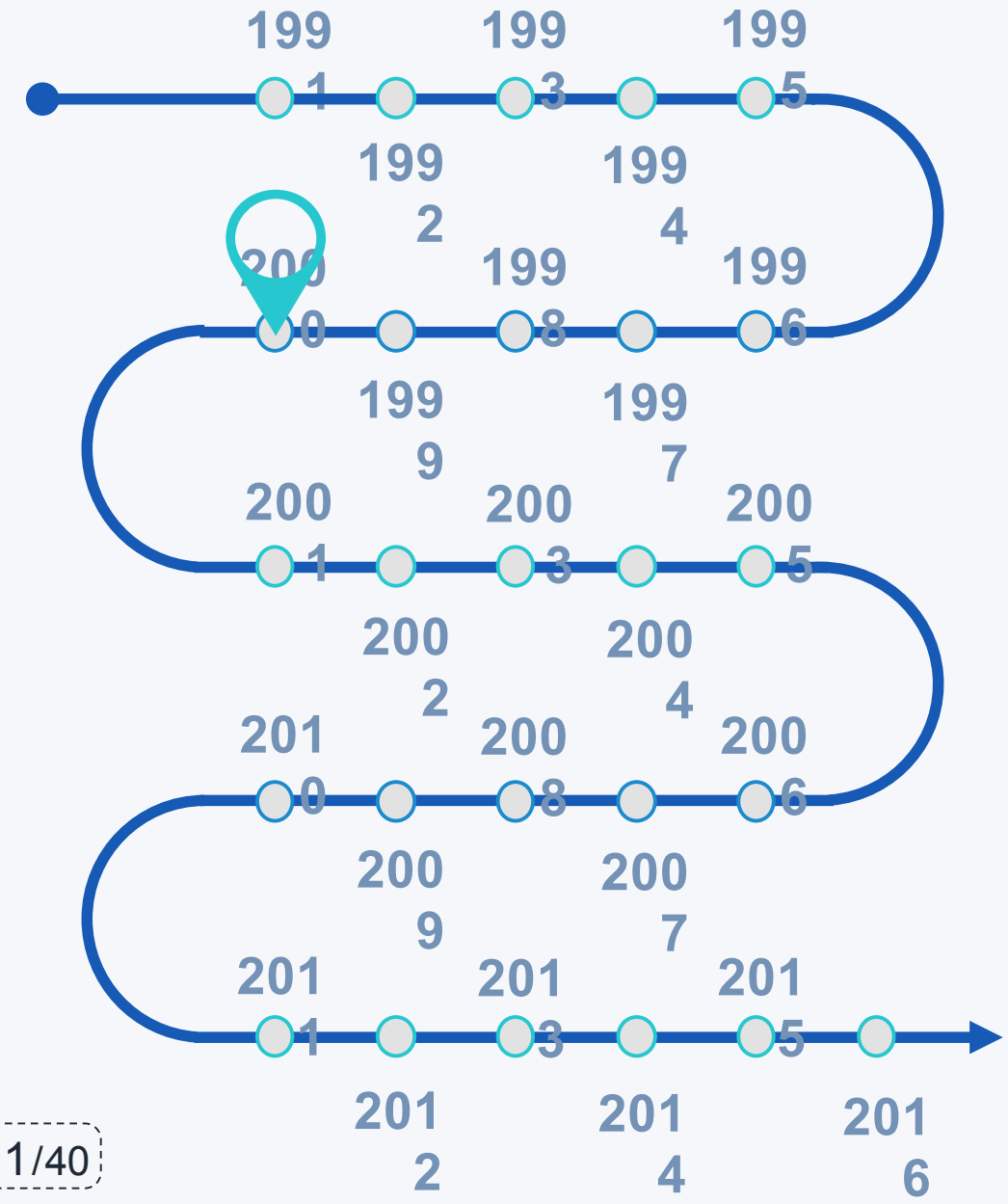
Imitation

T. Yamaguchi, M. Miura, and M. Yachida, “Multi-agent reinforcement learning with adaptive mimetism,” in *Proceedings 1996 IEEE Conference on Emerging Technologies and Factory Automation. ETFA ’96*, 1996, vol. 1, pp. 288–294.

حافظه جمعی

CM

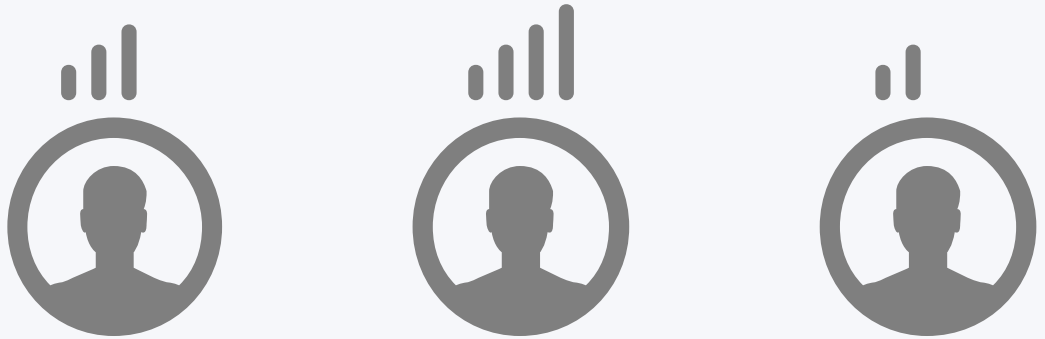
A. Garland and R. Alterman, “Multiagent learning through collective memory,” in *Adaptation, Coevolution and Learning in Multiagent Systems: Papers from the 1996 AAAI Spring Symposium*, 1996, pp. 33–38.

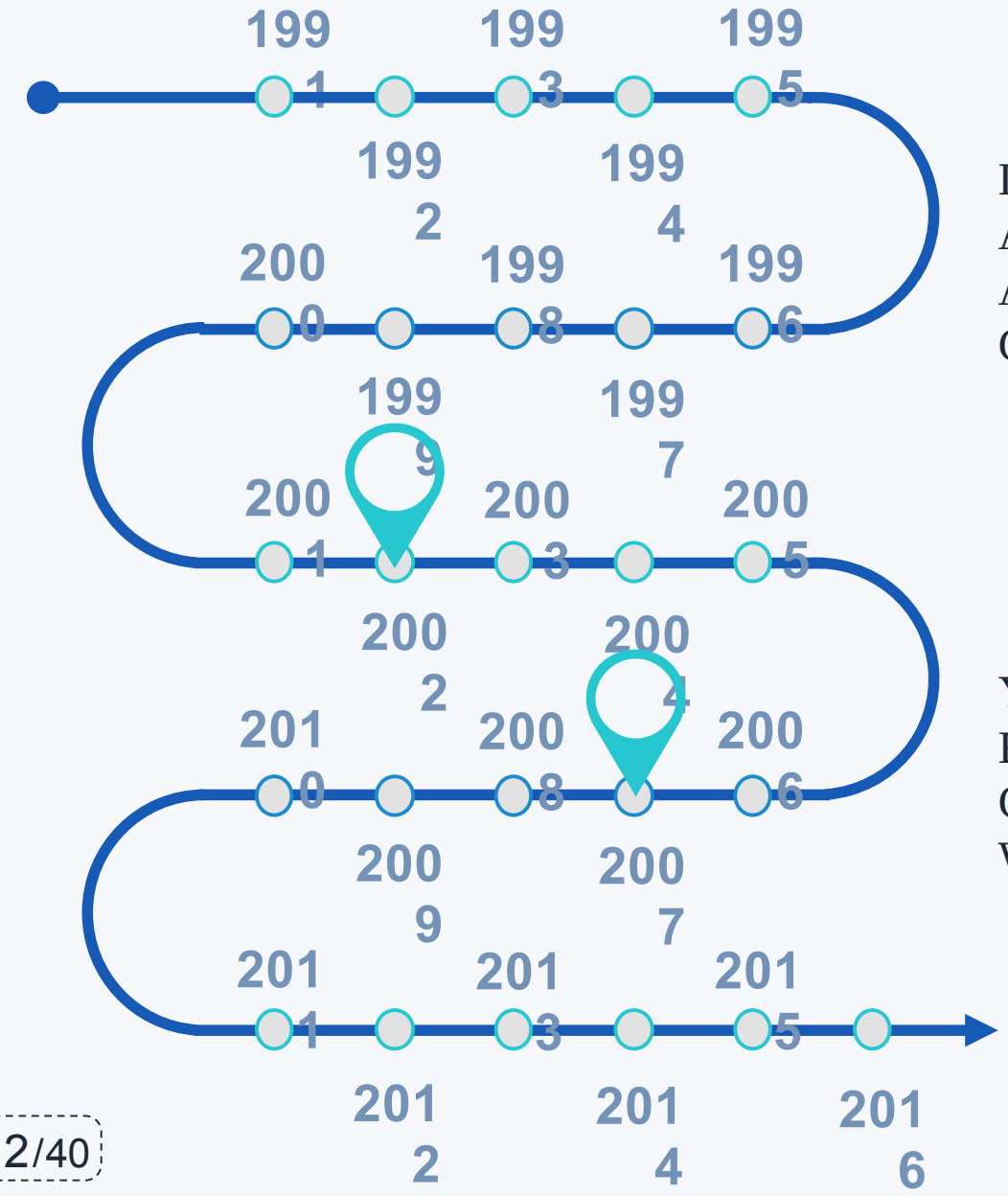


خبرگی

WSS

M. N. Ahmadabadi, M. Asadpur, S. H. Khodanbakhsh, and E. Nakano, "Expertness measuring in cooperative learning," in *Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000) (Cat. No.00CH37113)*, 2000, vol. 3, pp. 2261–2267.





پند

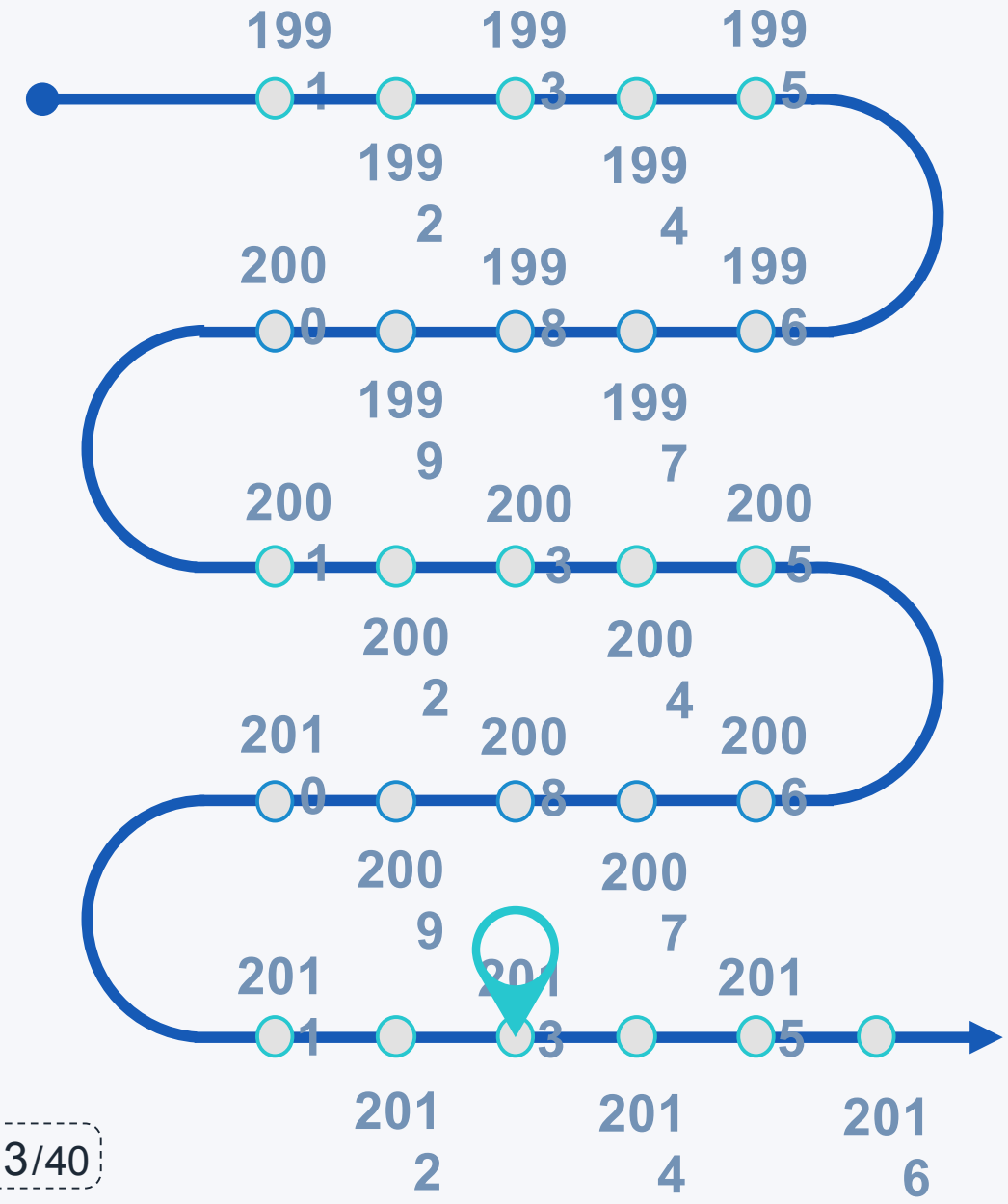
L. Nunes and E. Oliveira, “On Learning by Exchanging Advice,” Proc. Artif. Intell. Simul. Behav. Conv. Symp. Adapt. agents multi-agent Syst. (AISB/AAMAS-II), Imp. Coll. london, vol. cs.LG/0203, pp. 583–599

Advice

مبتنی بر پختگی سیاست

Y. Yang, Y. Tian, and H. Mei, “Cooperative Q Learning Based on Blackboard Architecture,” in 2007 International Conference on Computational Intelligence and Security Workshops (CISW 2007), 2007, pp. 224–227.

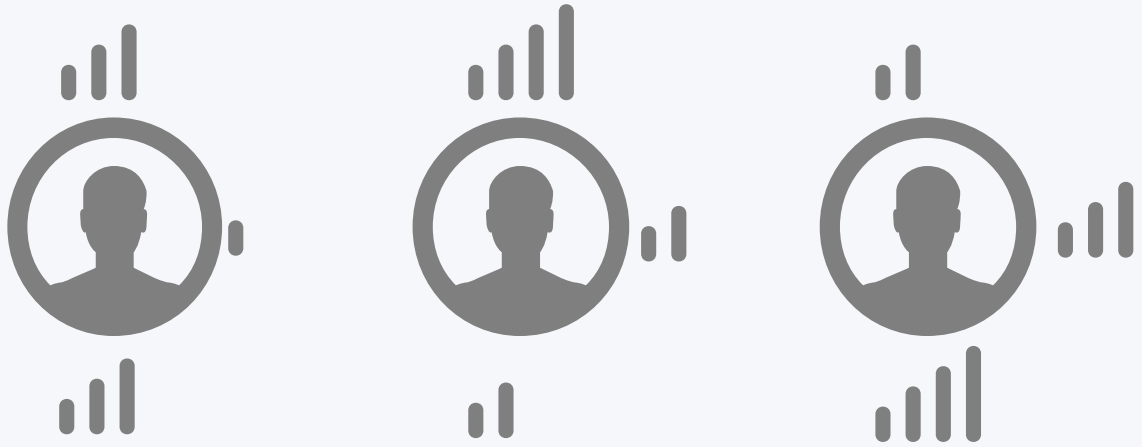
Blackboard



خبرگی چند معیاره

MCE

E. Pakizeh, M. Palhang, and M. M. Pedram, "Multi-criteria expertness based cooperative Q-learning," *Appl. Intell.*, vol. 39, no. 1, pp. 28–40, Jul. 2013.



1: معمولی

در نظر گرفتن پاداش و جریمه

2: مثبت

در نظر گرفتن پاداش

3: منفی

در نظر گرفتن جریمه

4: قدرمطلق

در نظر گرفتن جریمه و پاداش

به عنوان دو پارامتر مثبت

5: گرادیان

در نظر گرفتن پاداش تنها

در آخرین چرخه یادگیری

6: میانگین فاصله

ارزیابی بر اساس میانگین

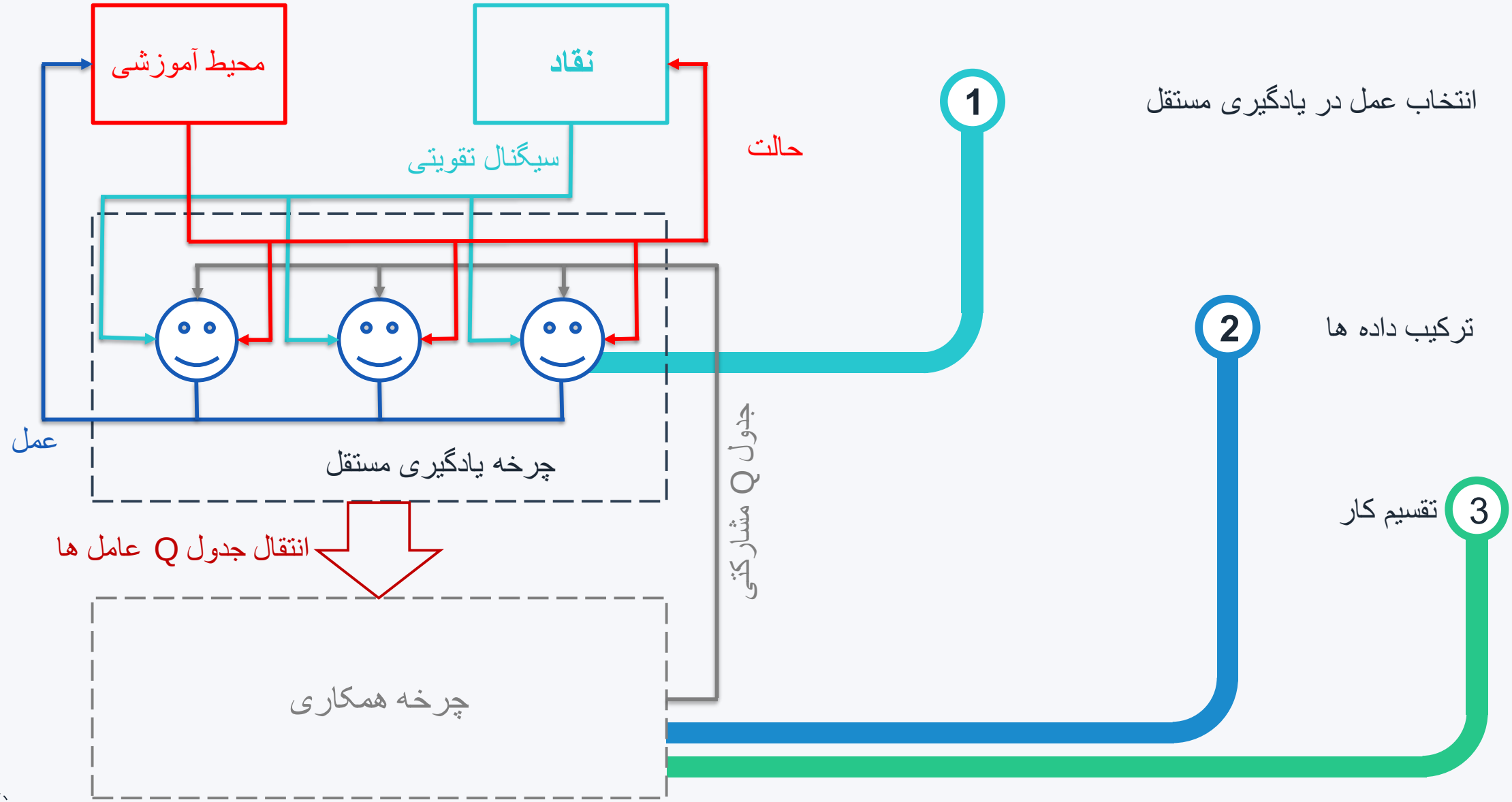
تعداد قدم های برداشته

برای رسیدن به هدف

تشریح

روش

پیشنهادی


محیط آموزشی
نقاد
سیگنال تقویتی
حالت
عمل
چرخه یادگیری مستقل
انتقال جدول Q عامل ها
چرخه همکاری
جدول Q مشارکتی
1
انتخاب عمل در یادگیری مستقل
2
ترکیب داده ها
3
تقسیم کار

معیار ارزیابی ارائه شده

عنوان: شوک

محاسبه براساس: تعداد دفعاتی که پاداش یکی از

خانه های جذب در یک خانه از ماتریس Q اثر داشته است.

معیار ارزیابی ارائه شده

↑	←	↓	→	
0	0	0	0	1
0	0	0	0	2
0	0	0	8	3
0	0	0	0	4
0	0	0	0	5
0	0	0	0	6
0	0	0	0	7
0	0	0	0	8
0	0	0	0	9

جدول Q

↑	←	↓	→	
0	0	0	1	1
0	0	0	2	2
0	1	1	1	3
0	0	0	0	4
1	0	1	0	5
0	0	0	0	6
0	0	0	0	7
1	0	0	0	8
0	0	0	0	9

جدول مشاهده

↑	←	↓	→	
0	0	0	0	1
0	0	0	0	2
0	0	0	1	3
0	0	0	0	4
0	0	0	0	5
0	0	0	0	6
0	0	0	0	7
0	0	0	0	8
0	0	0	0	9

جدول شوک

1	2	3	4
		5	
6	7	8	9

معیار ارزیابی ارائه شده

عنوان: کوتاهترین فاصله تجربه شده

محاسبه براساس: کوتاهترین طول مسیر تا هر یک

از اهداف وبا انتخاب هر عمل که مشاهده شده است

معیار ارزیابی ارائه شده

عنوان: کوتاهترین فاصله تجربه شده

محاسبه براساس: کوتاهترین طول مسیر تا هر یک

از اهداف وبا انتخاب هر عمل که مشاهده شده است

4	3	2	1	
∞	∞	∞	3	1
∞	∞	∞	2	2
∞	∞	∞	1	3
∞	∞	∞	∞	4
2	∞	∞	∞	5
∞	∞	∞	∞	6
∞	∞	∞	∞	7
3	∞	∞	∞	8
∞	∞	∞	∞	9

2	3	4	0	3	0	0	5	0
1	1	1	0	4	0	0	4	0
0	0	0	0	0	0	0	0	0
1	2	3	4	5	6	7	8	9

انتخاب عمل در یادگیری مستقل

$$\pi(s)_t = \text{Boltzmann}(Q_t, \tau_1)$$

$$p_i = \frac{e^{-\varepsilon_i/kT}}{\sum_{j=1}^{M} e^{-\varepsilon_j/kT}}$$

$$\pi(s)_t = (1-\mu)\ \text{Boltzmann}(Q_t, \tau_1) + \ \mu\ \text{Boltzmann}(\frac{1}{\text{SEP}_t}, \tau_2)$$

تقسیم کار

$\pi(SEP) == \pi(\mathrm{Q})$

$\pi(SEP) != \pi(\mathrm{Q})$

$\pi(SEP) != \pi(\mathrm{Q})$

$\pi(SEP) == \pi(\mathrm{Q})$

$\pi(SEP) == \pi(\mathrm{Q})$

$Group_1 =$

$Group_2 =$

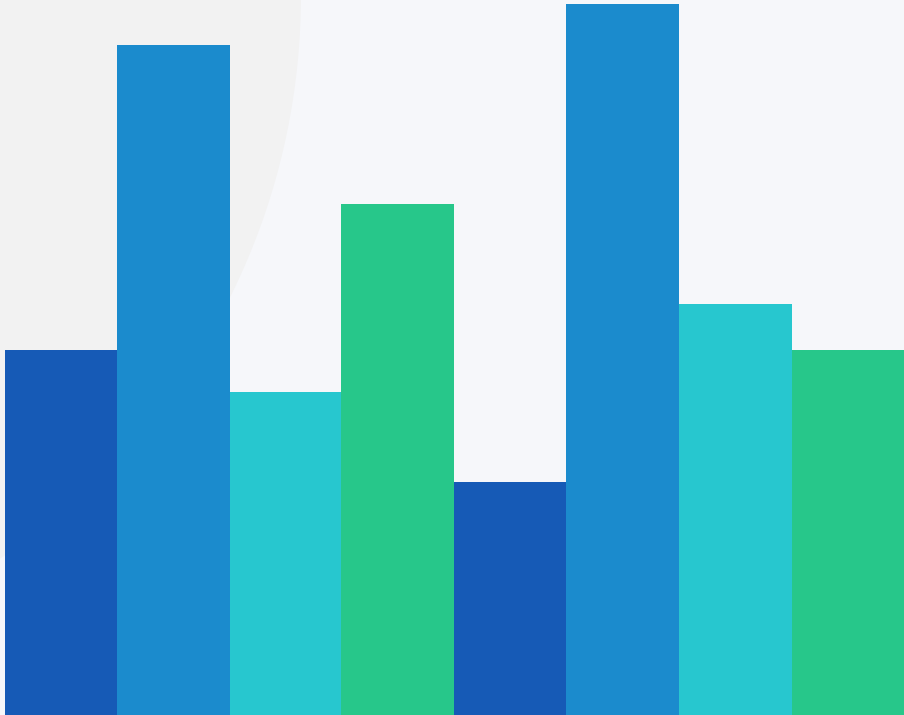
ترکیب داده ها

$$Q_{co}(s) = \sum_{i \in G} \frac{\sum_{k \in actionSet} shock_i(s,k)}{\sum_{j \in G} \sum_{k \in actionSet} shock_i(s,k)} Q_i\ (s)$$

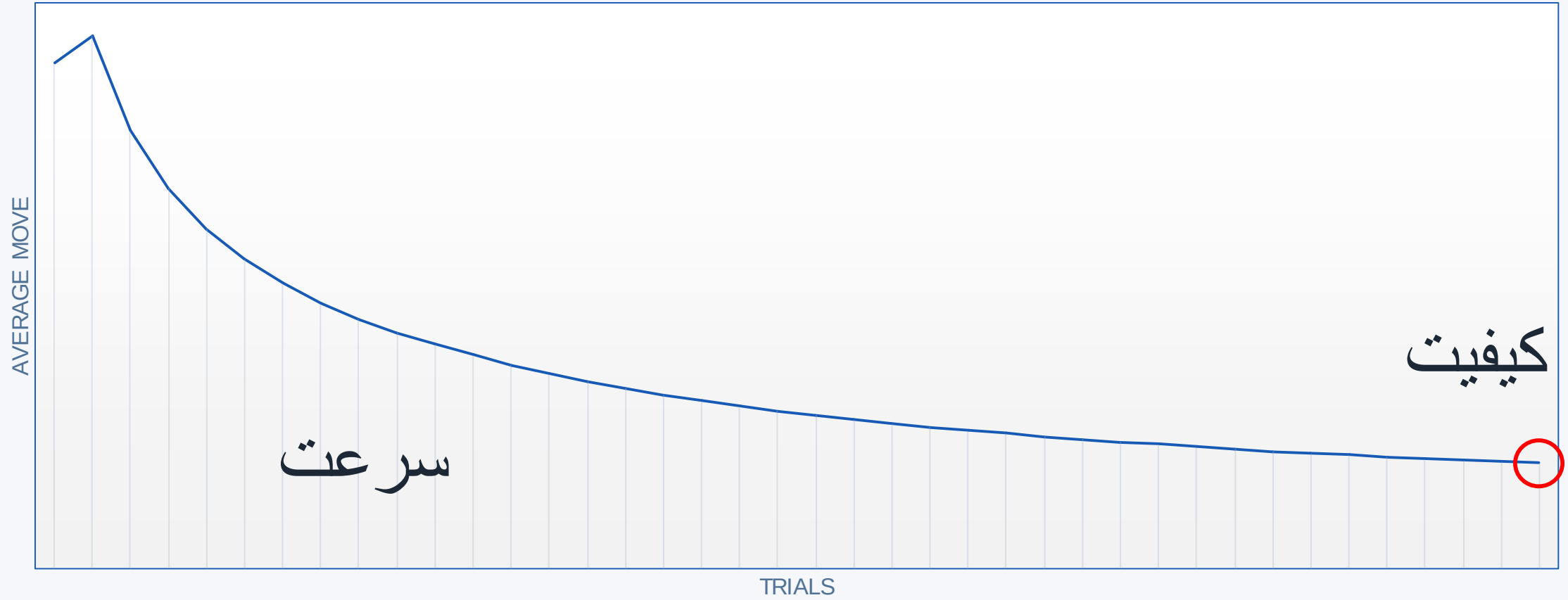
SEP

نتیجه

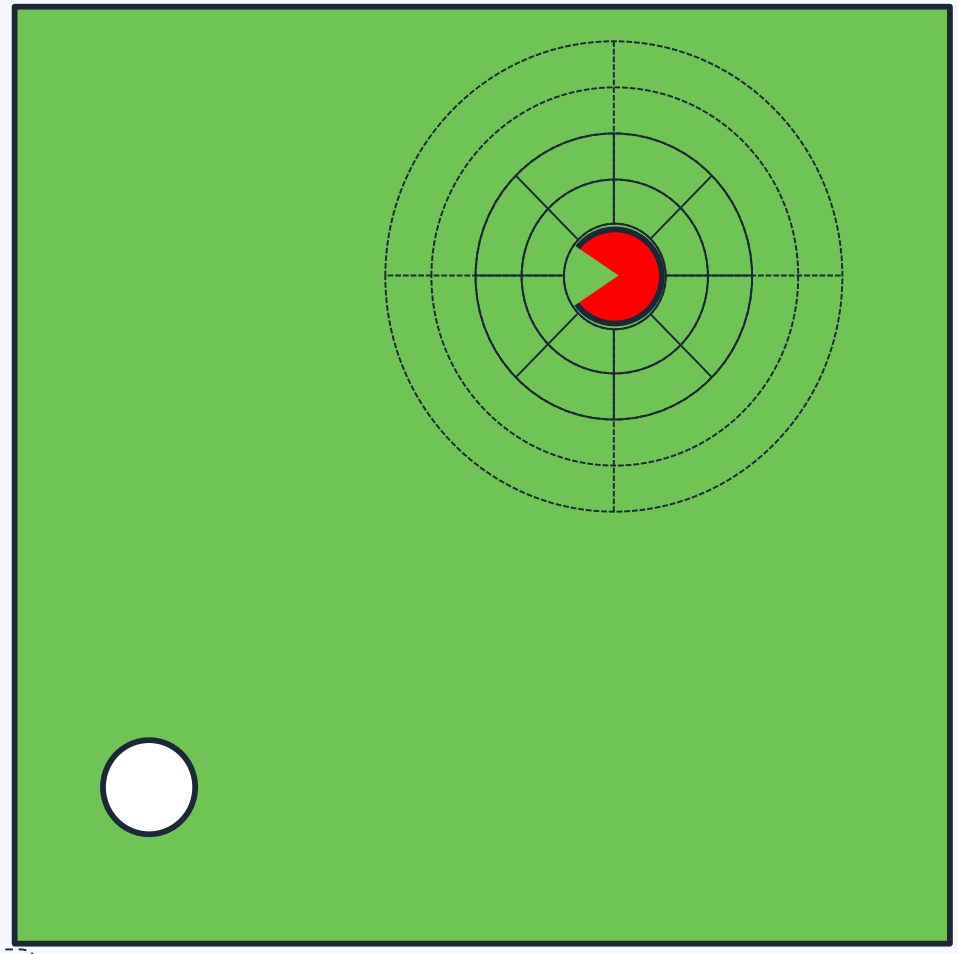
آزمایشات



بهره گیری از کوتاهترین فاصله تجربه شده در یادگیری مستقل



صید و صیاد



16 عمل

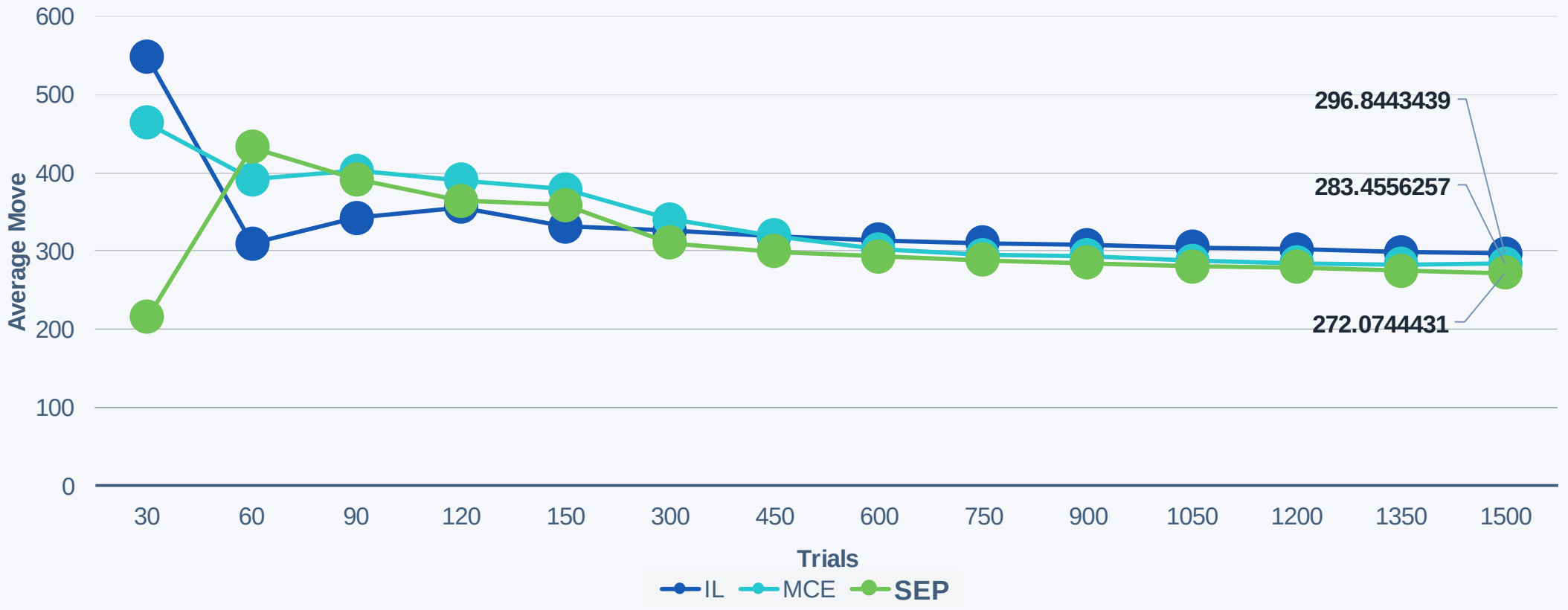
16 حالت

پاداش هر شکار 10 و جریمه 0.1 برای هر حرکت بی حاصل

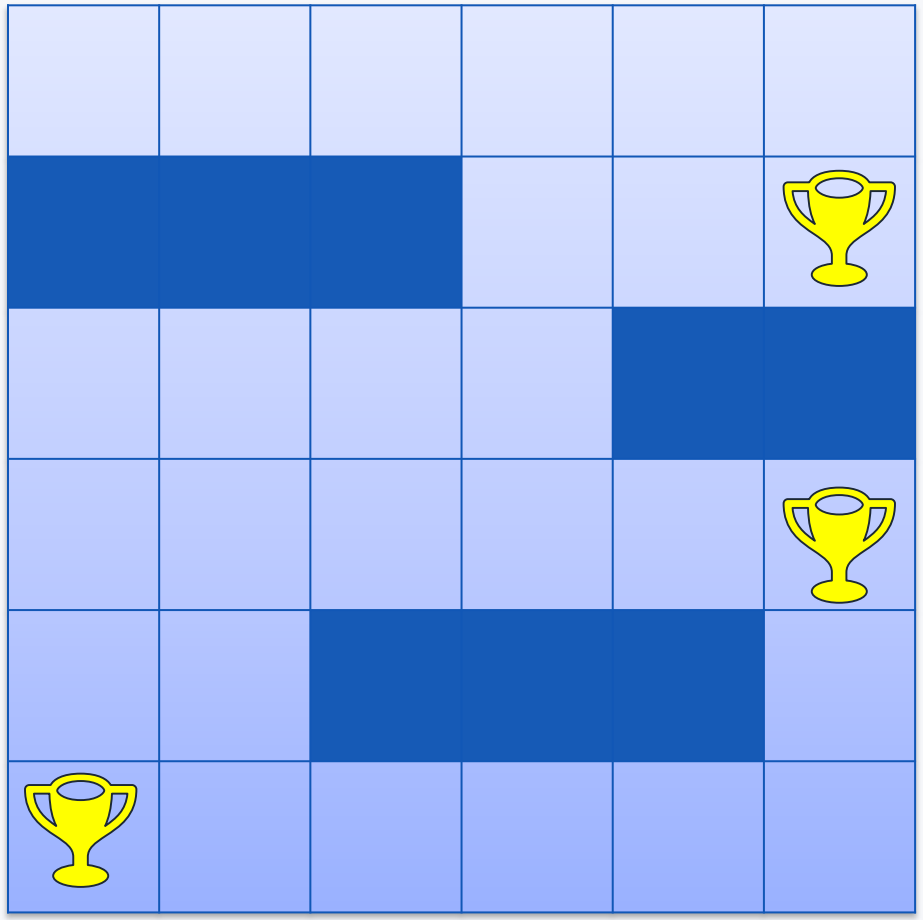
صیاد صید

بهره گیری از روش پیشنهادی در مقایسه با روش خبرگی چند معیاره

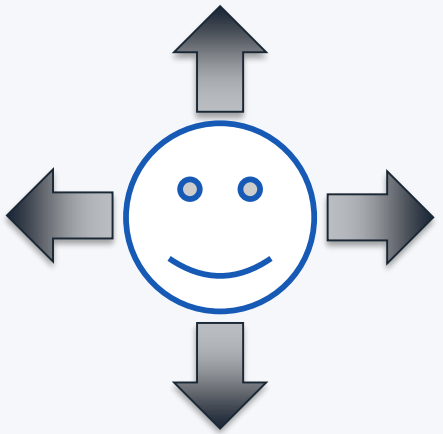
محیط صید و صیاد



پلکان مارپیچ



28حالت



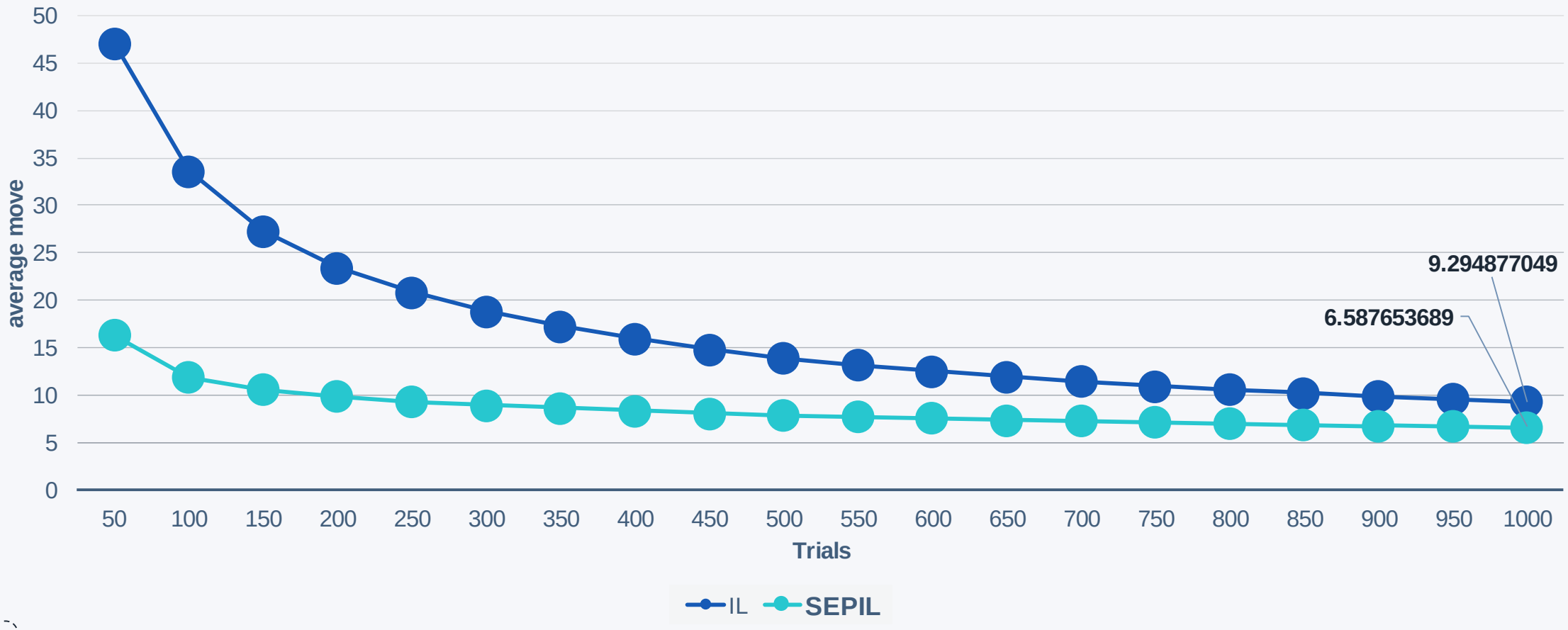
4حالت

پاداش کشف طلا 10 و جریمه بر خورد یا موانع 1- پاداش دیگر حرکت ها بر

اساس رابطه زیر

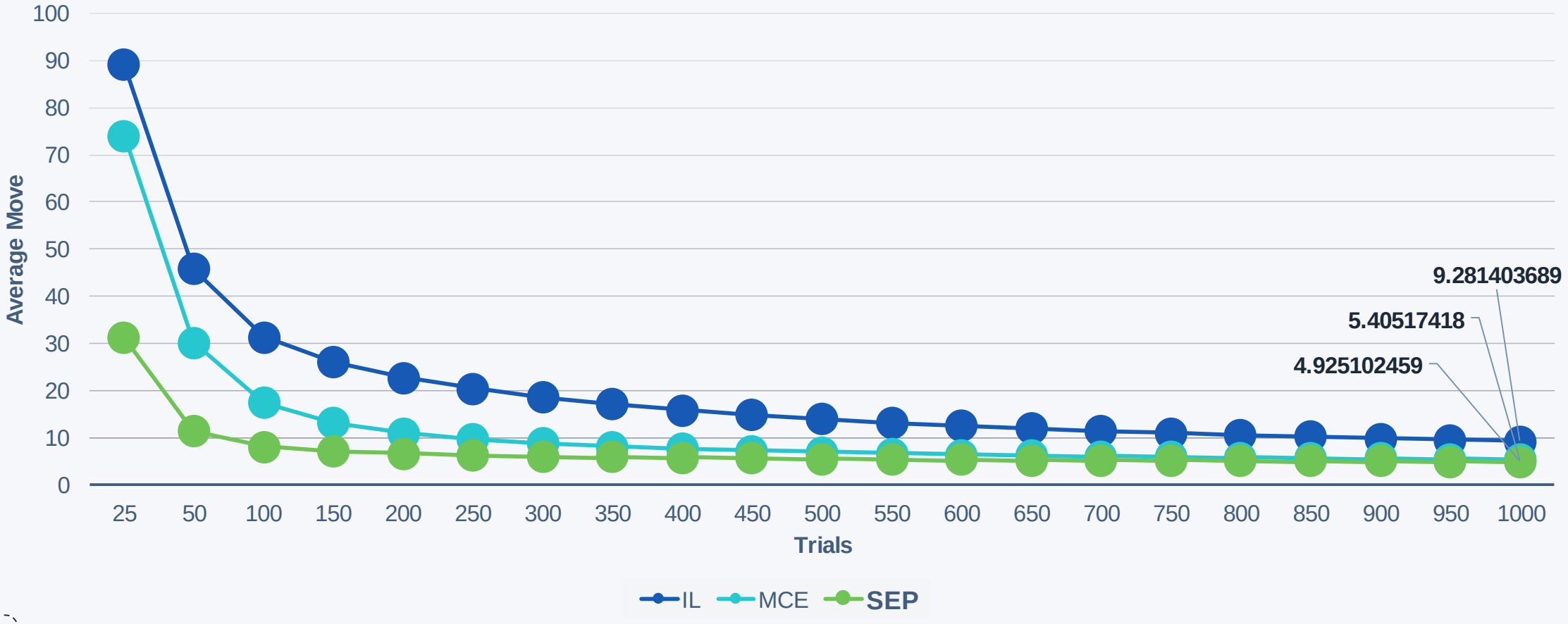
$$Reward = \frac{1}{distance\ between\ the\ agent\ and\ the\ goal}$$

بهره گیری از کوتاهترین فاصله تجربه شده در یادگیری مستقل

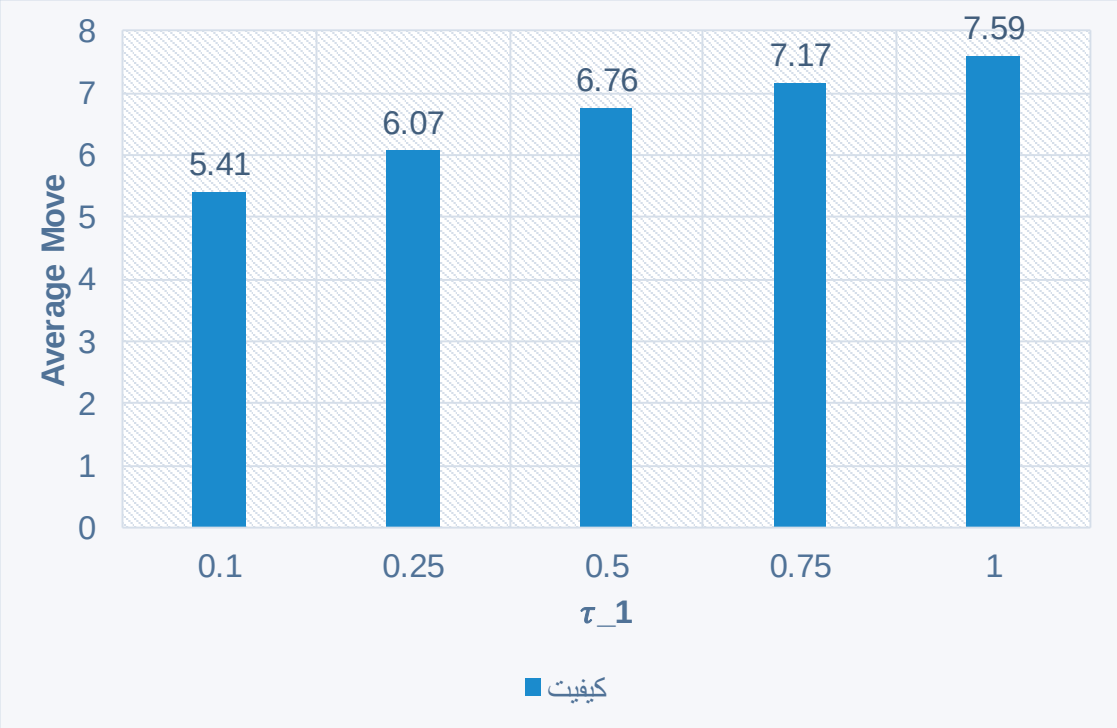


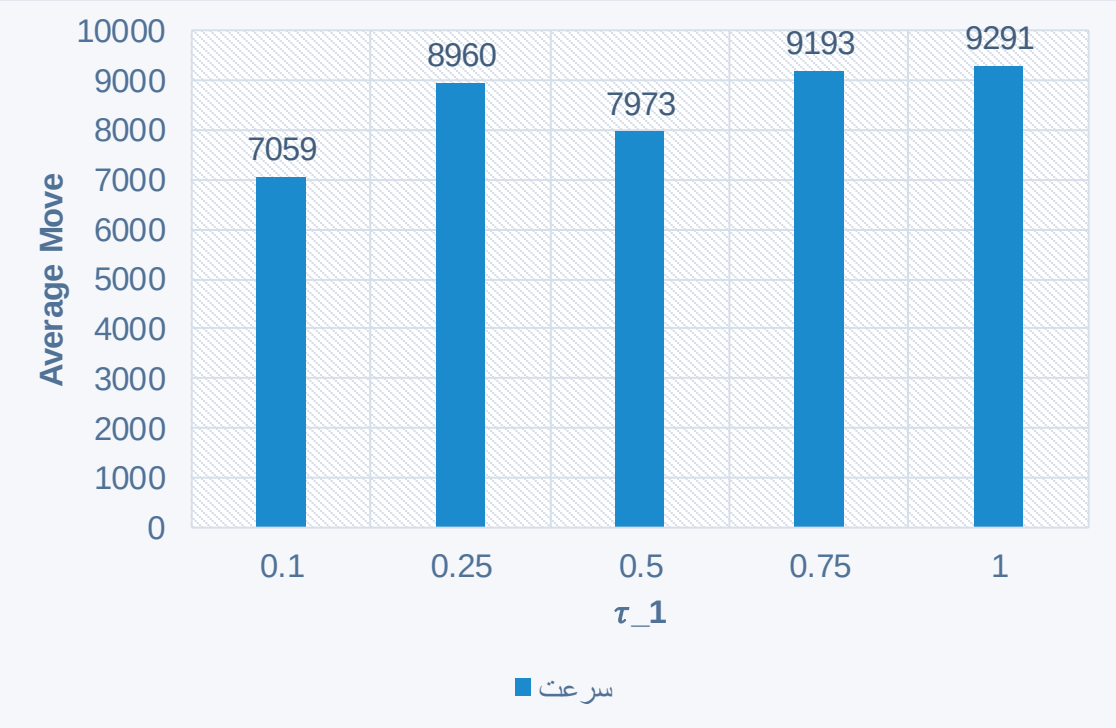
بهره گیری از روش پیشنهادی در مقایسه با روش خبرگی چند معیاره

محیط پلکان مارپیچ

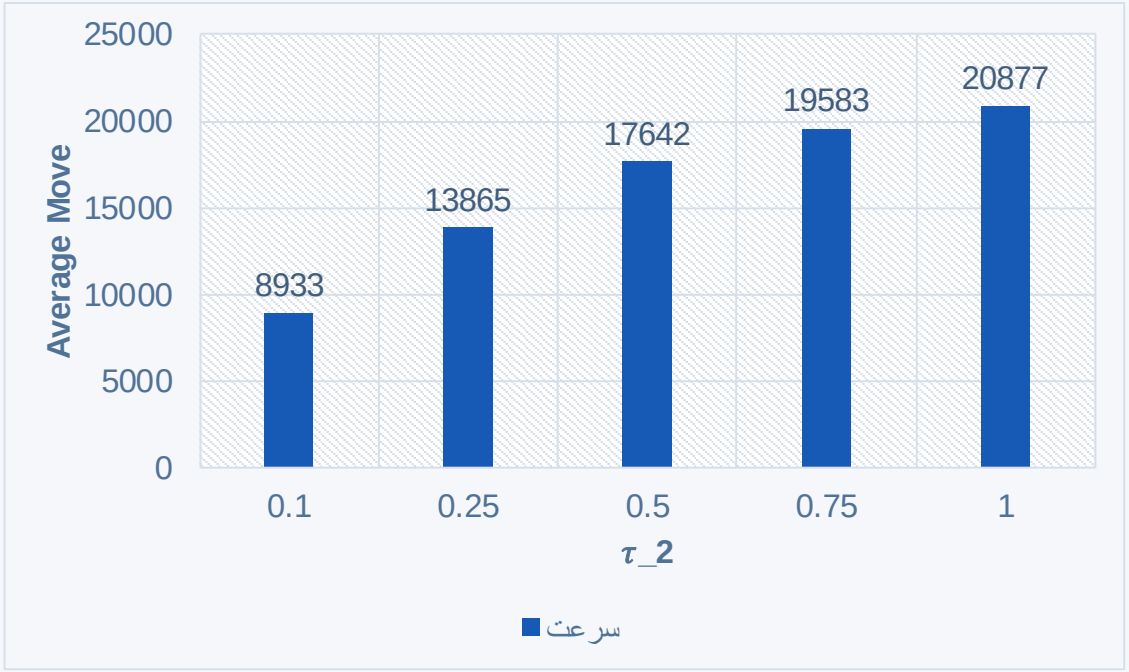


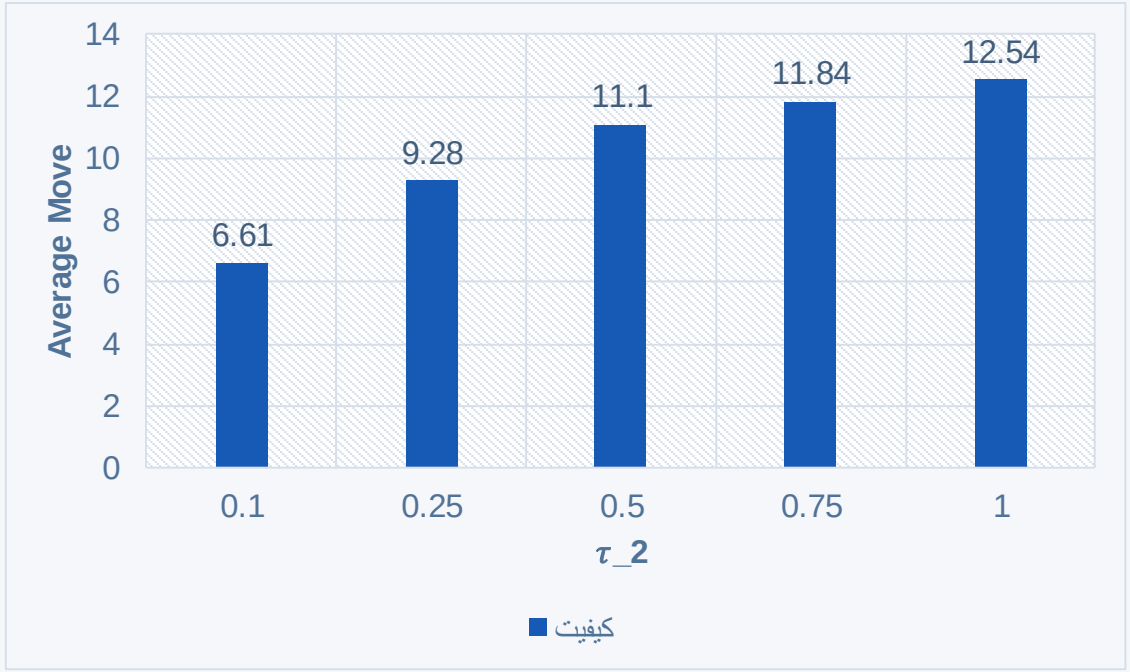
بررسی حساسیت روش پیشنهادی در برابر پارامتر τ_1



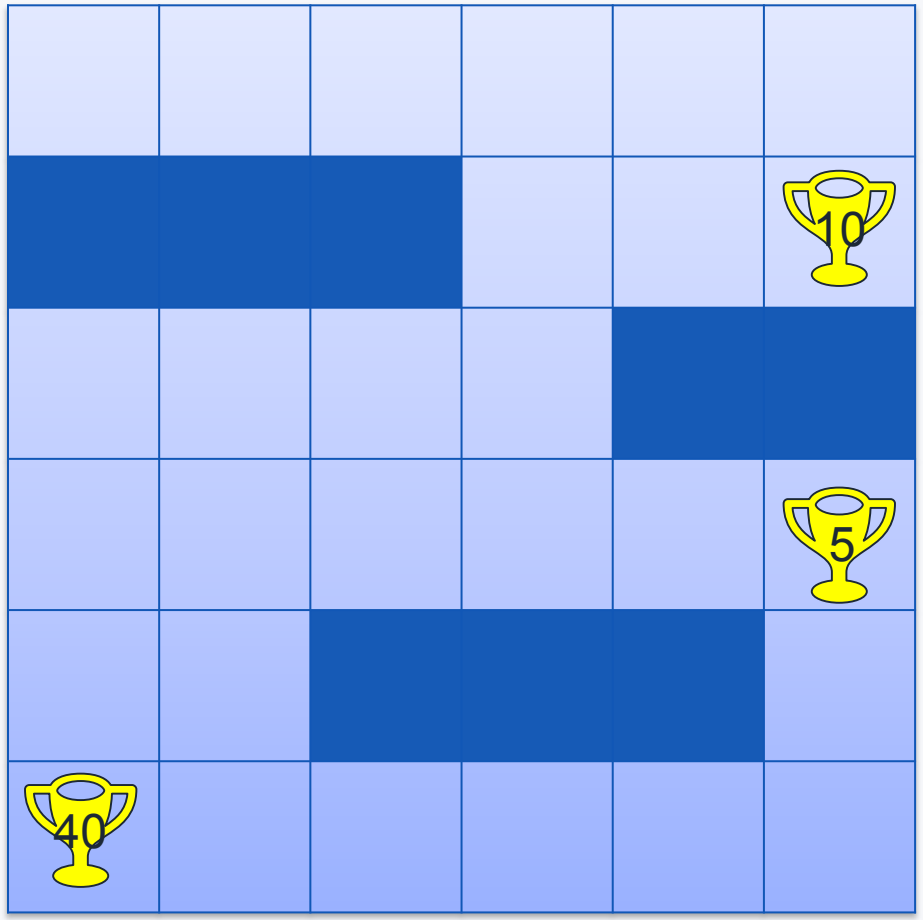


بررسی حساسیت روش پیشنهادی در برابر پارامتر τ_2

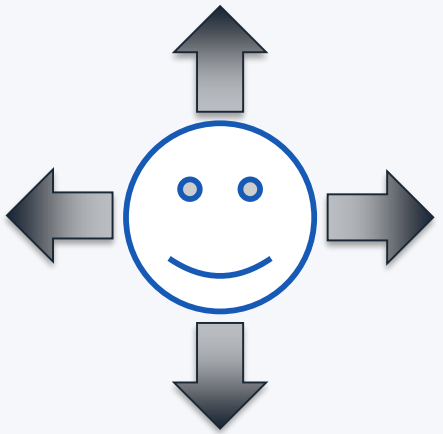




پلکان مارپیچ



28حالت

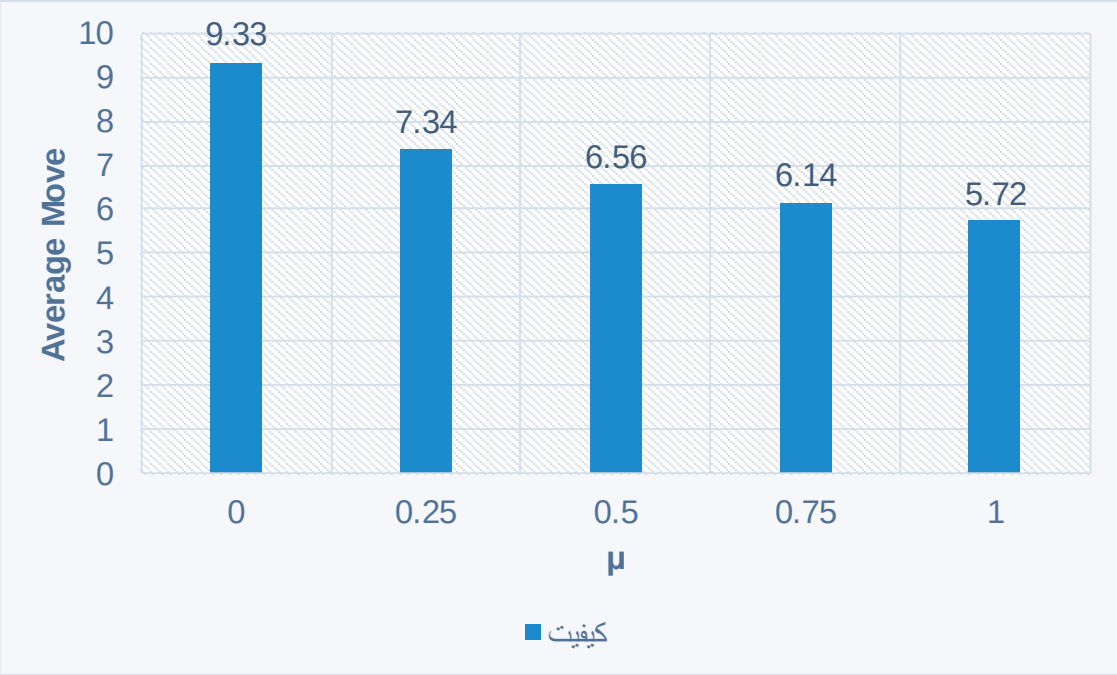


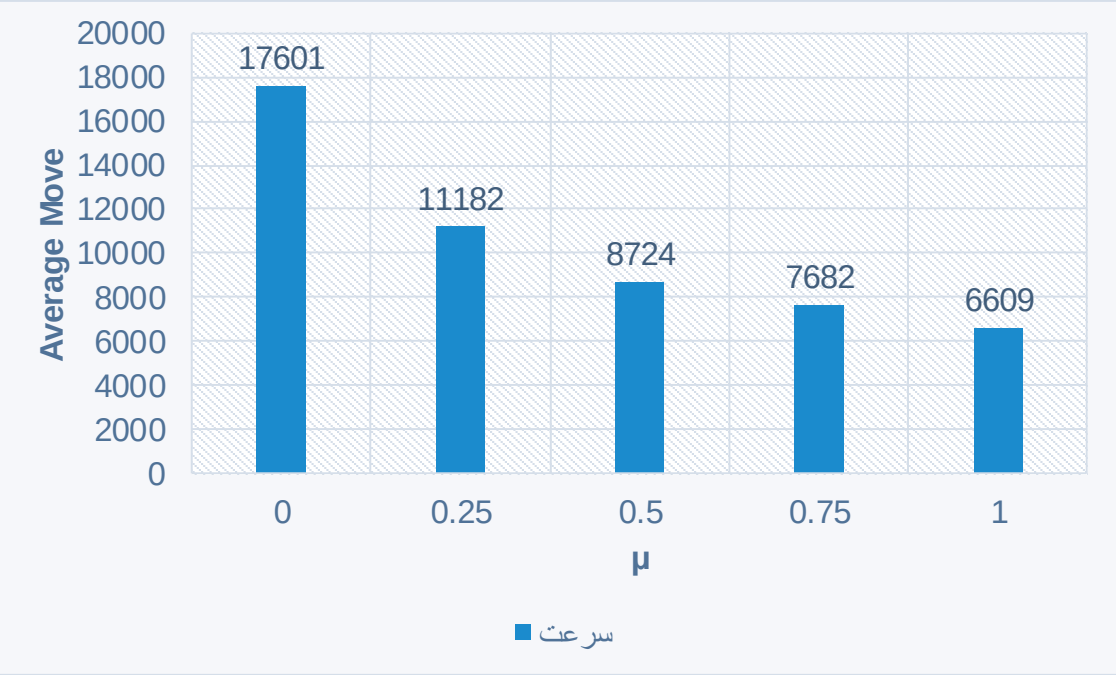
4عمل

جریمه بر خورد با موانع 1- پاداش دیگر حرکت ها بر اساس رابطه زیر

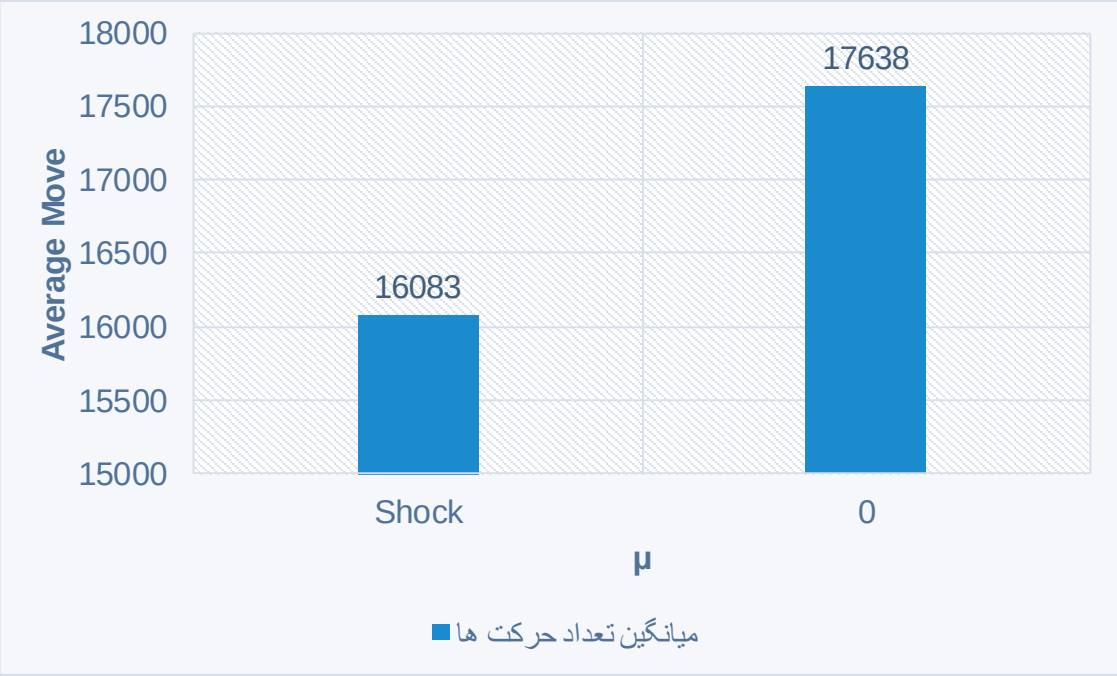
$$Reward = \frac{1}{distance\ between\ the\ agent\ and\ the\ goal}$$

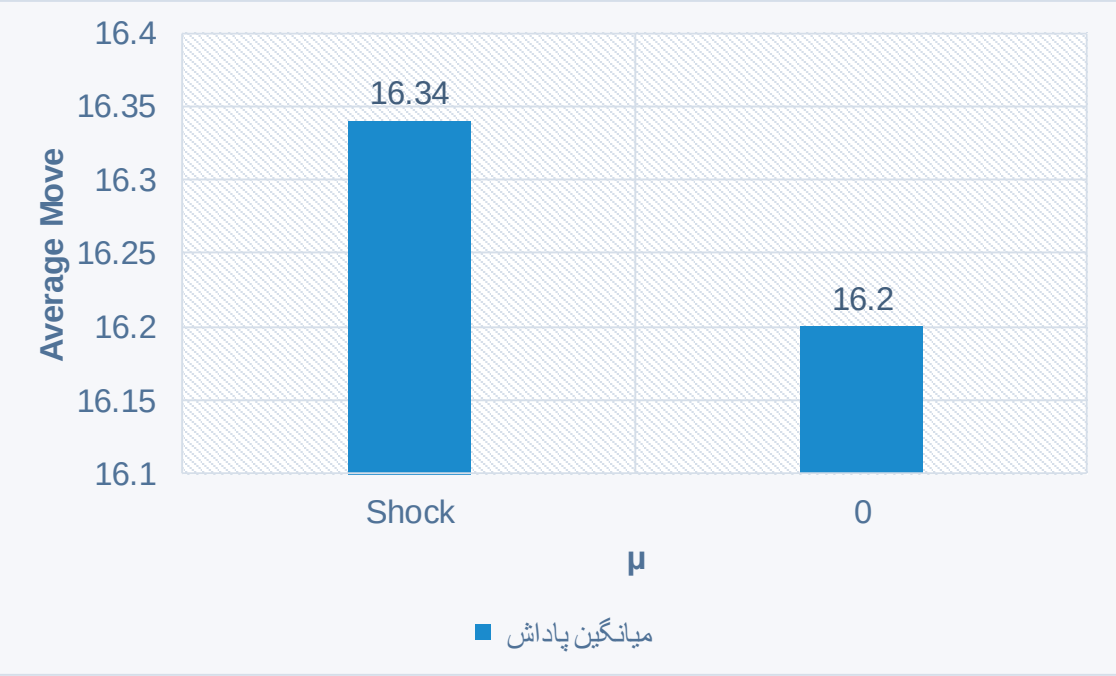
بررسی حساسیت روش پیشنهادی در برابر پارامترμ





بررسی حساسیت روش پیشنهادی در برابر پارامترμ





جمع‌بندی

جمع بندی

- ارائه معیار هایی جهت ارزیابی عامل های یادگیری مشارکتی
- بهره گیری از معیار های ارائه شده در راستای تسریع یادگیری مشارکتی
- بهبود یادگیری تقویتی با بهره گیری از اطلاعات تجربی

پیشنهادات

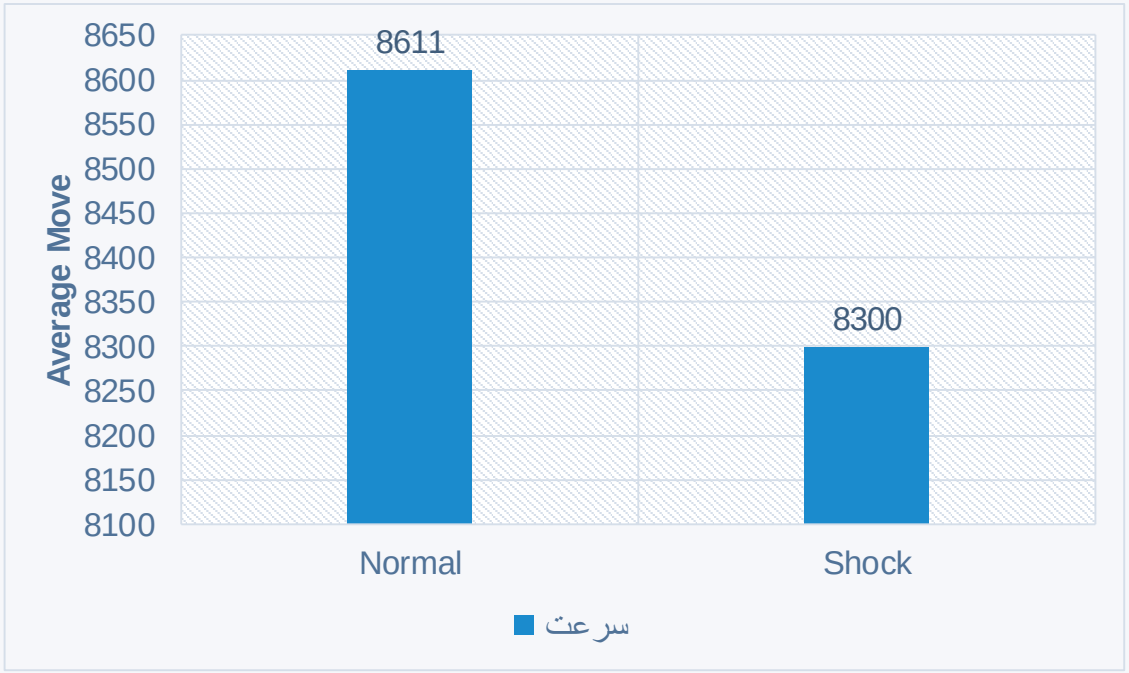
- متغیر کردن میزان بهره گیری از حداقل فاصله تجربه شده.
- تقسیم کار مناسب.
- تهیه معیارهایی مشابه معیار کوتاهترین فاصله تجربه شده.

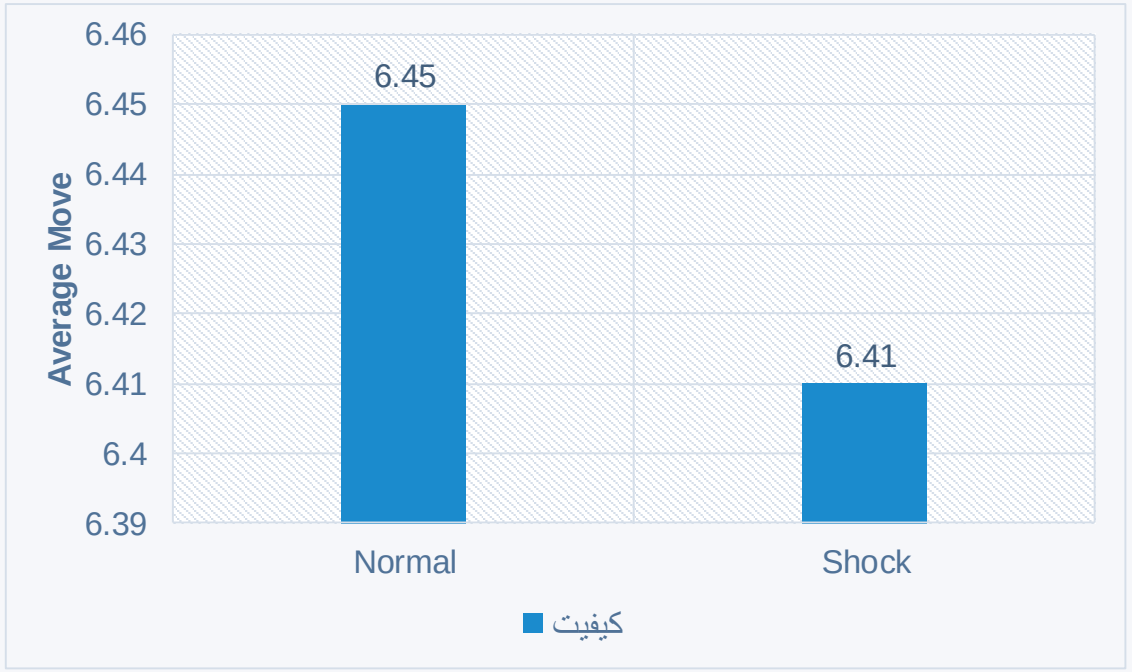
باتشکر از

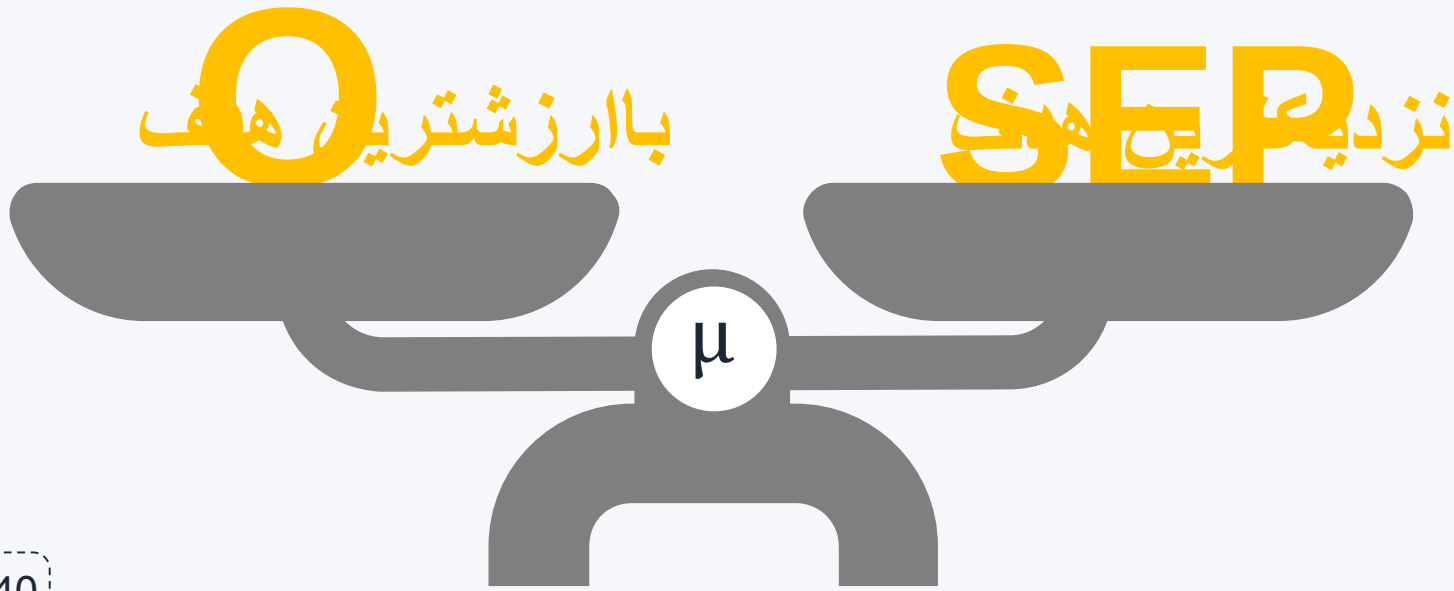
توجه شما

پرسش و پاسخ

تاثیر استفاده از شوک

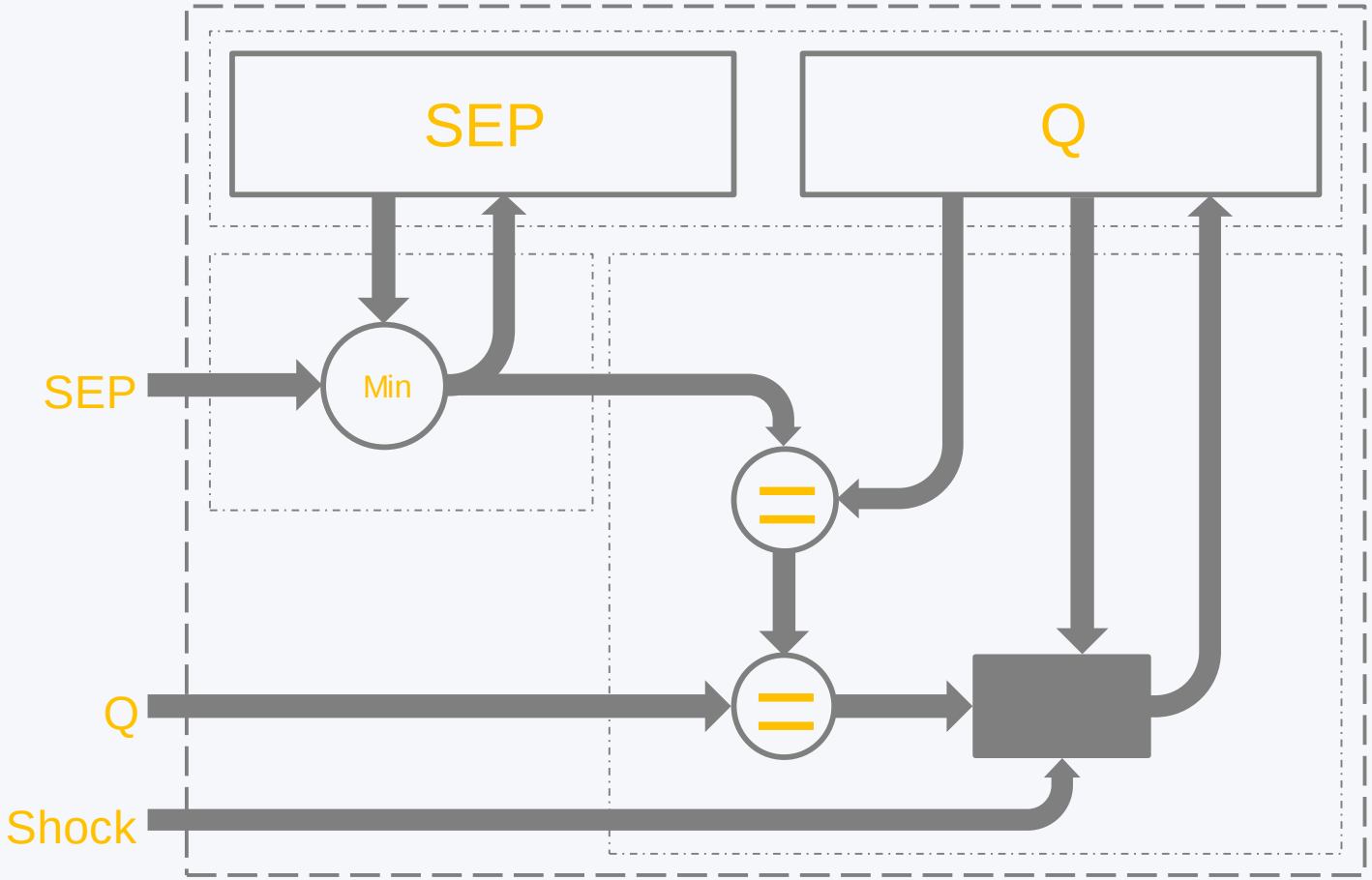





O
باارزشترین هدف
SEP
نزدیک‌ترین همسایه
μ


thank you!

ترکیب داده ها



محیط های آزمایشی