Language Understanding for Text-based Games Using Deep Reinforcement Learning

Karthik Narasimhan, Tejas Kulkarni, Regina Barzilay

MIT

Text-based games

(State 1: The old bridge)

You are standing very close to the bridge's eastern foundation. If you go east you will be back on solid ground ... The bridge sways in the wind.

>> go east

(State 2: Ruined gatehouse)

The old gatehouse is near collapse. Part of its northern wall has already fallen down ... East of the gatehouse leads out to a small open area surrounded by the remains of the castle. ...

MUDs: predecessors to modern graphical games

Why are they challenging?

(State 1: The old bridge)

You are standing very close to the bridge's eastern foundation. If you go east you will be back on solid ground ... The bridge sways in the wind.

~~*Location:* Bridge 1~~

~~*Wind level:* 3~~

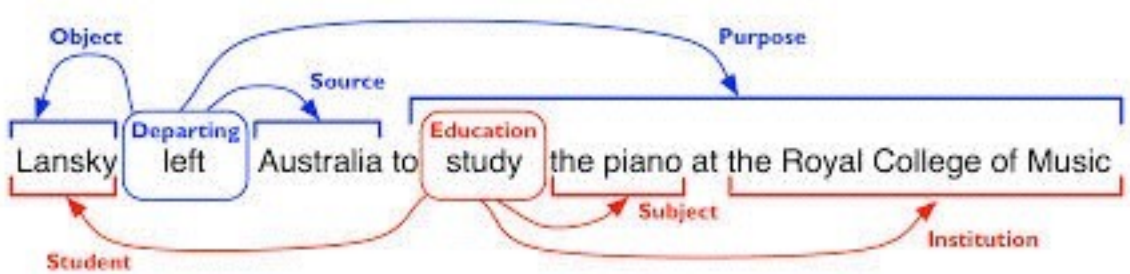
~~*Time*: 8pm~~

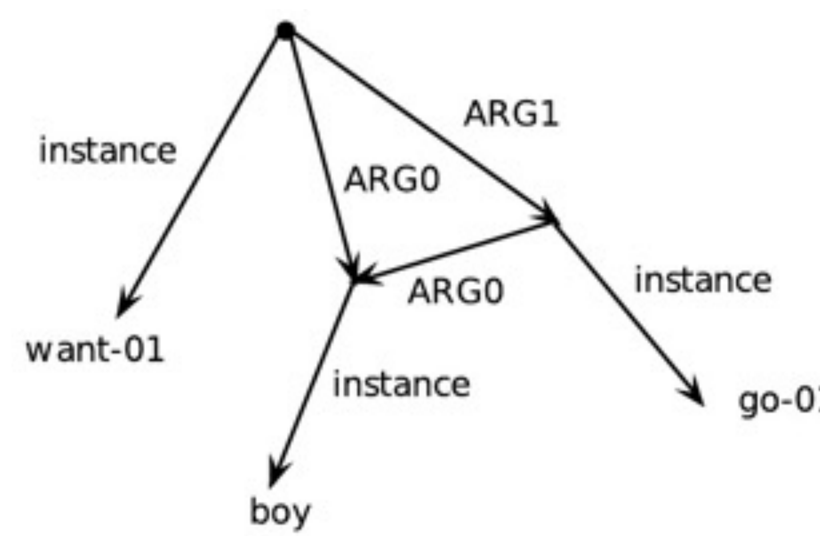
Branavan et al., 2011

No symbolic representation available

Can a computer understand language enough in order to play these games?

Understanding $\approx$ Actionable intelligence





Can a computer understand language enough in order to play these games?

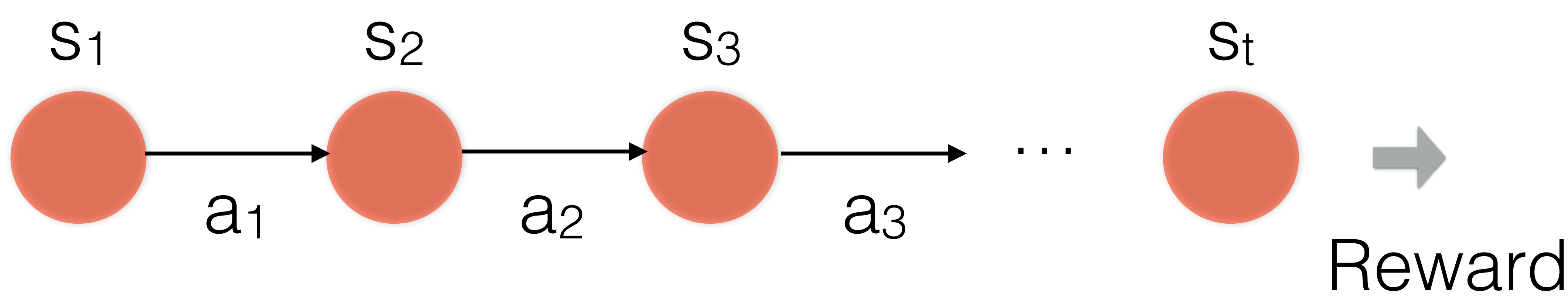
Inspiration: Playing graphical games directly from raw pixels (DeepMind)

Our Approach

Reinforcement Learning utilizing in-game feedback to:

- ✦ Learn control policies for gameplay.
- ✦ Learn good representations for text description of game state.

Traditional RL framework



s =

Location: Bridge 1

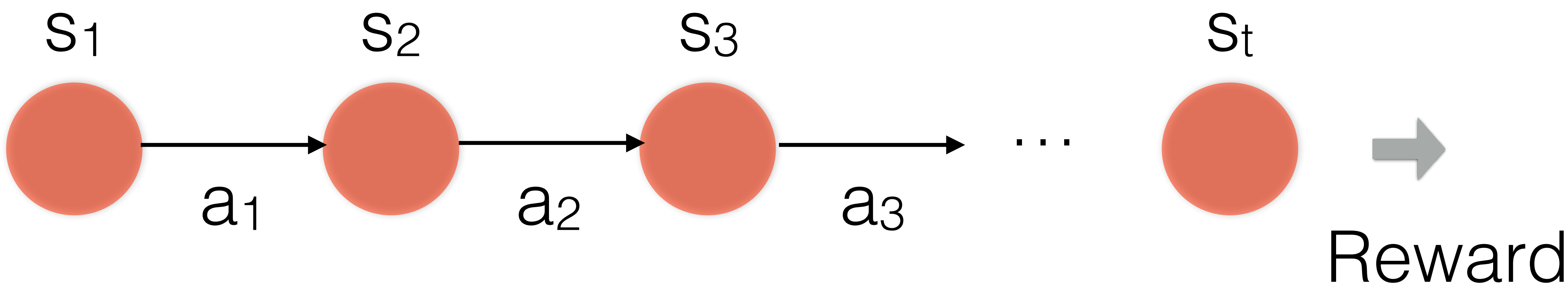
Wind level: 3

Time: 8pm

$$Q(s, a)$$

Q-value is the agent's notion of discounted future reward

Text-based games



$s =$

~~*Location:* Bridge 1~~

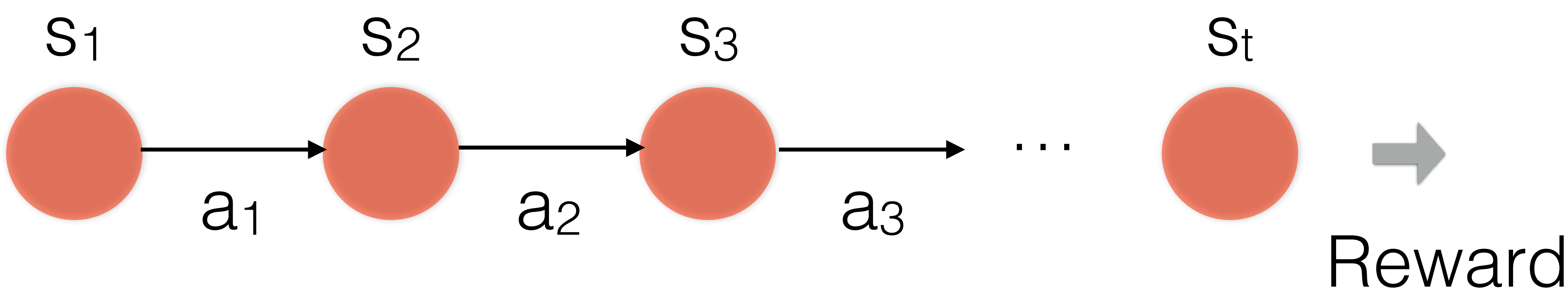
~~*Wind level:* 3~~

~~*Time*: 8pm~~

(State 1: The old bridge)

You are standing very close to the bridge's eastern foundation. If you go east you will be back on solid ground ...

Text-based games: BOW representation

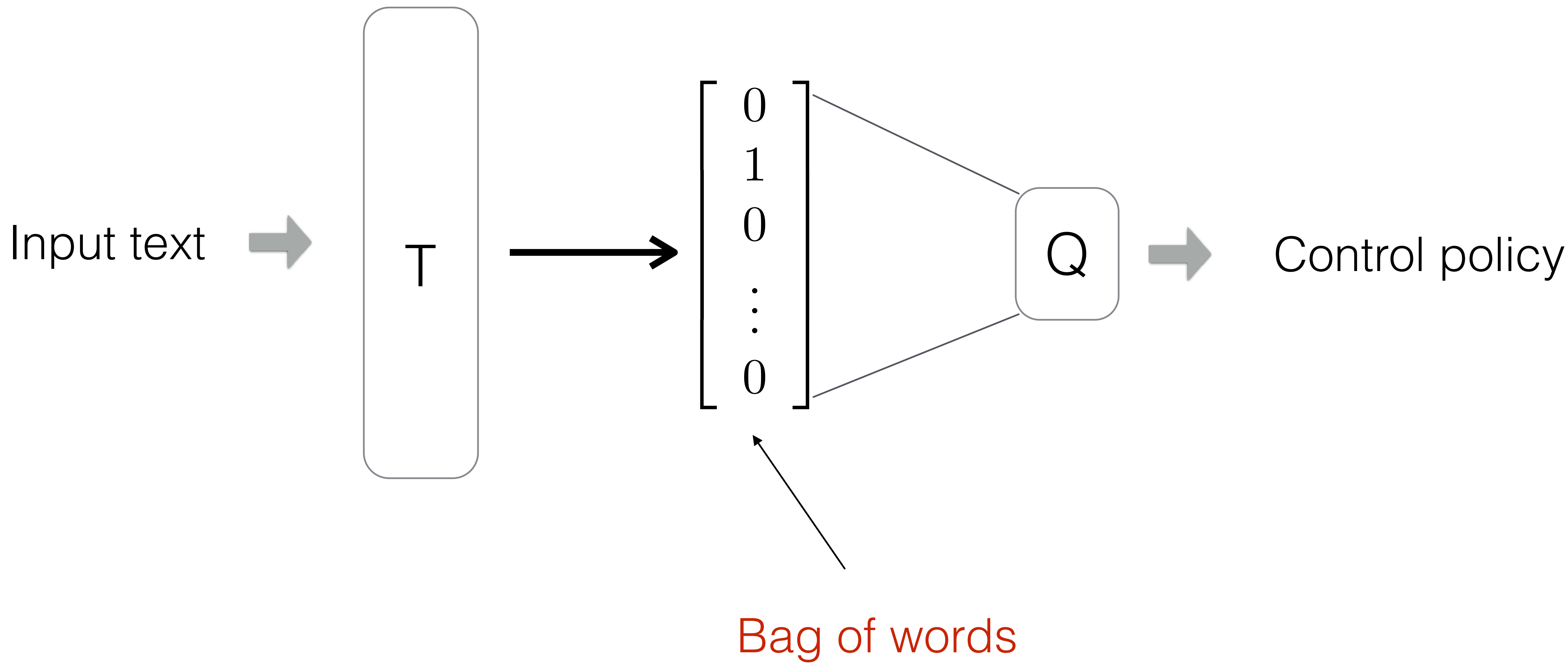


$$s = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Bag of words?

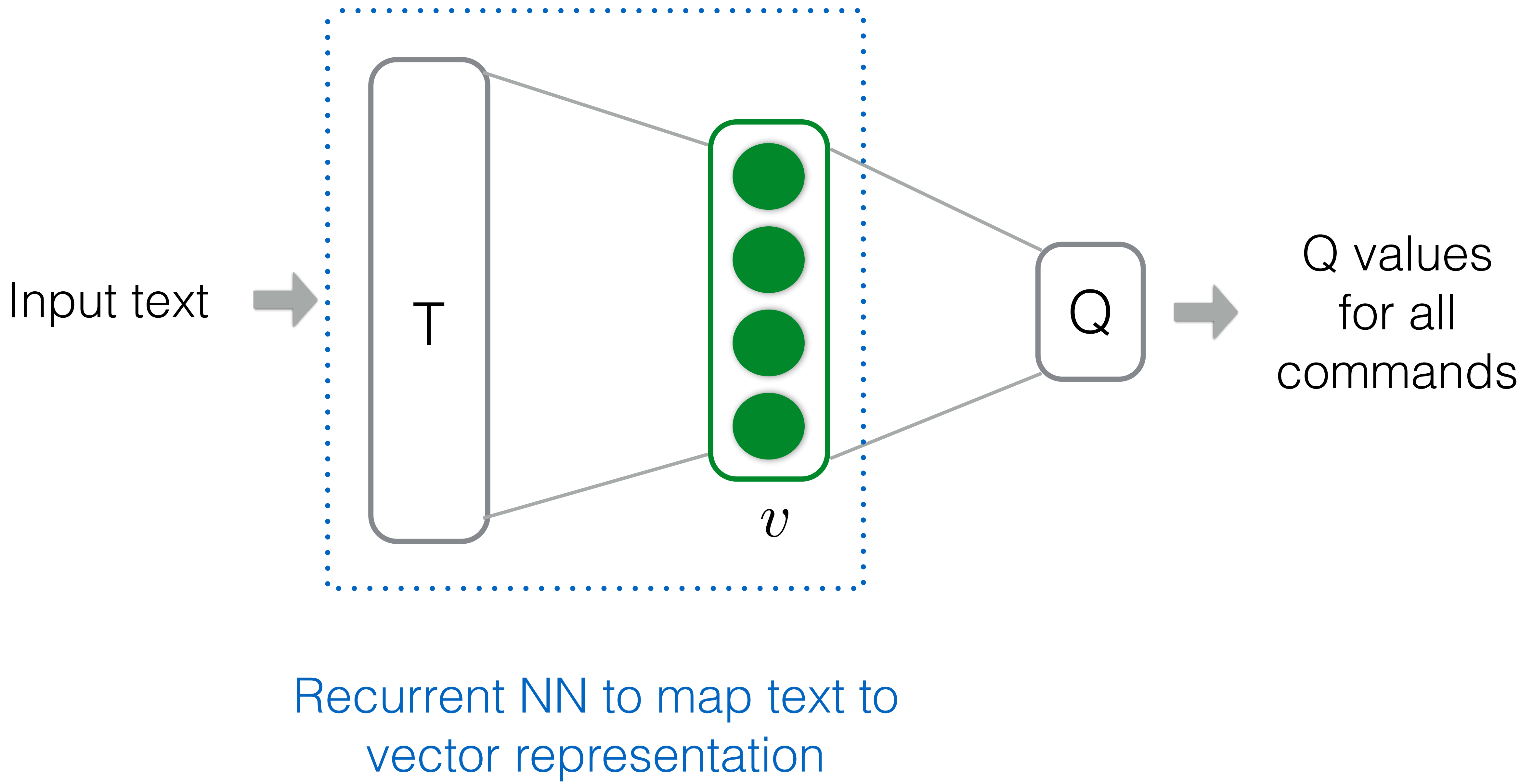
(State 1: The old bridge)

You are standing very close to the bridge's eastern foundation. If you go east you will be back on solid ground ...

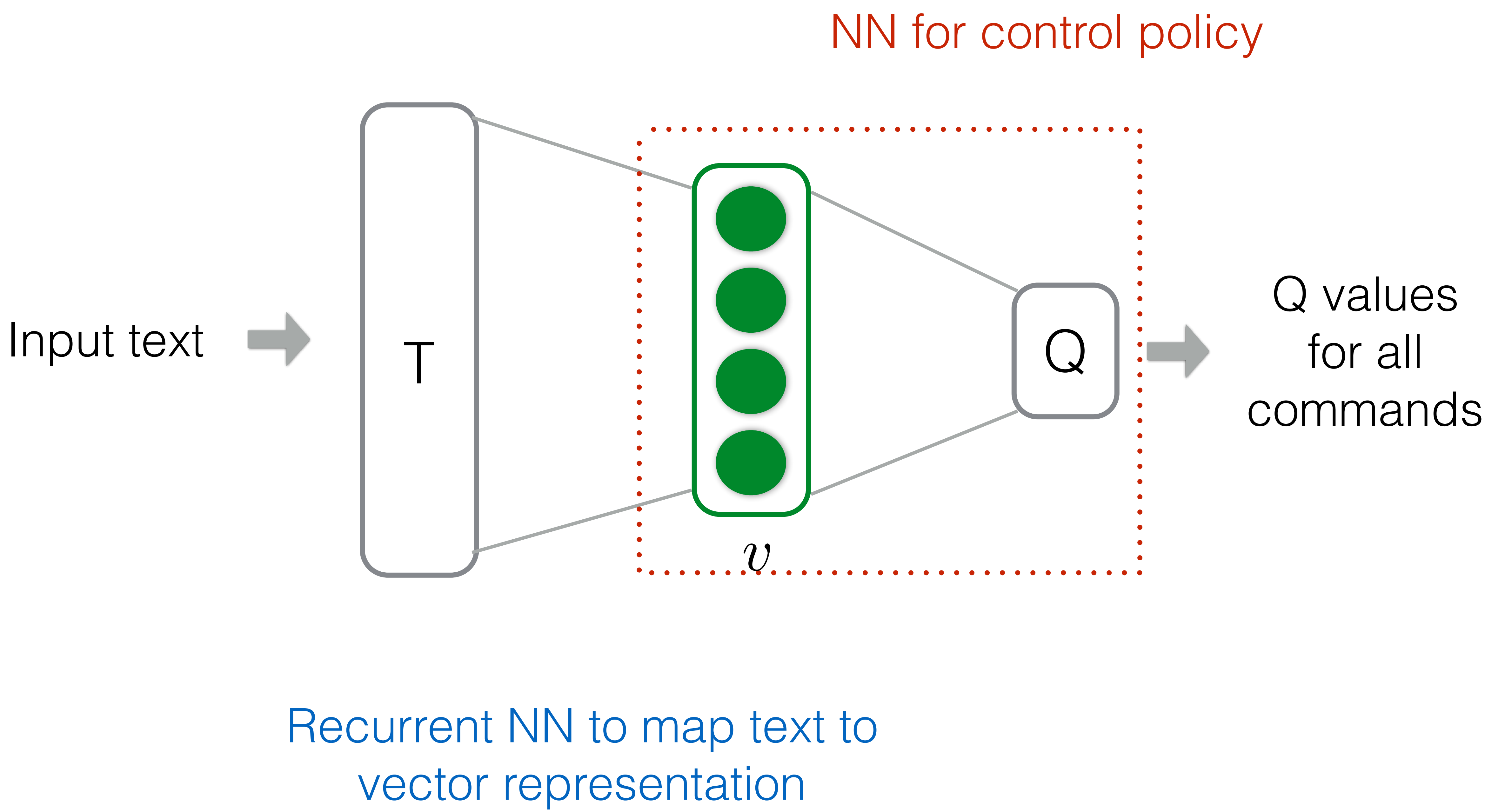


Can we do better?

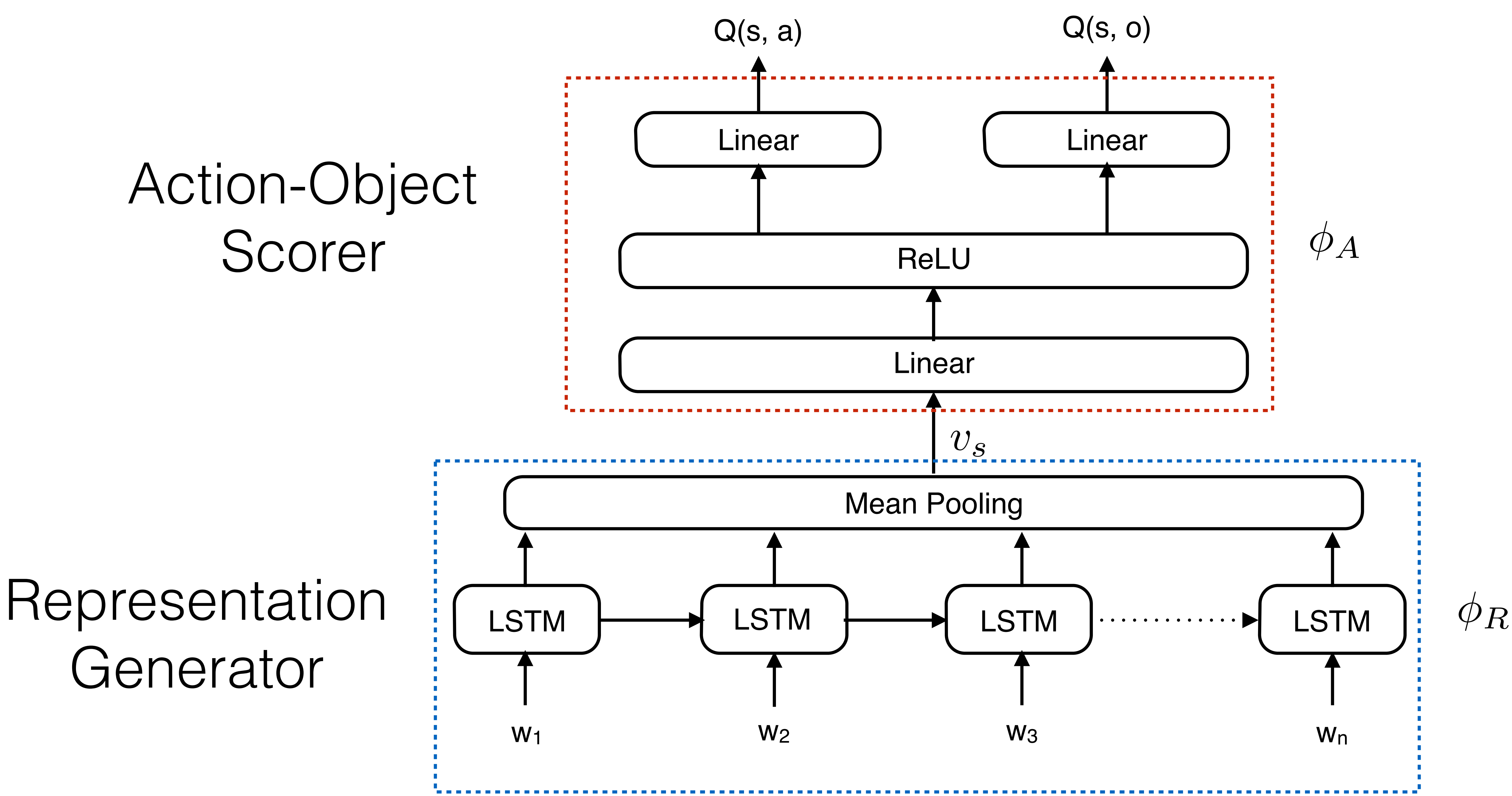
Model


Input text
T
v
Q
Q values
for all
commands
Recurrent NN to map text to
vector representation

Model

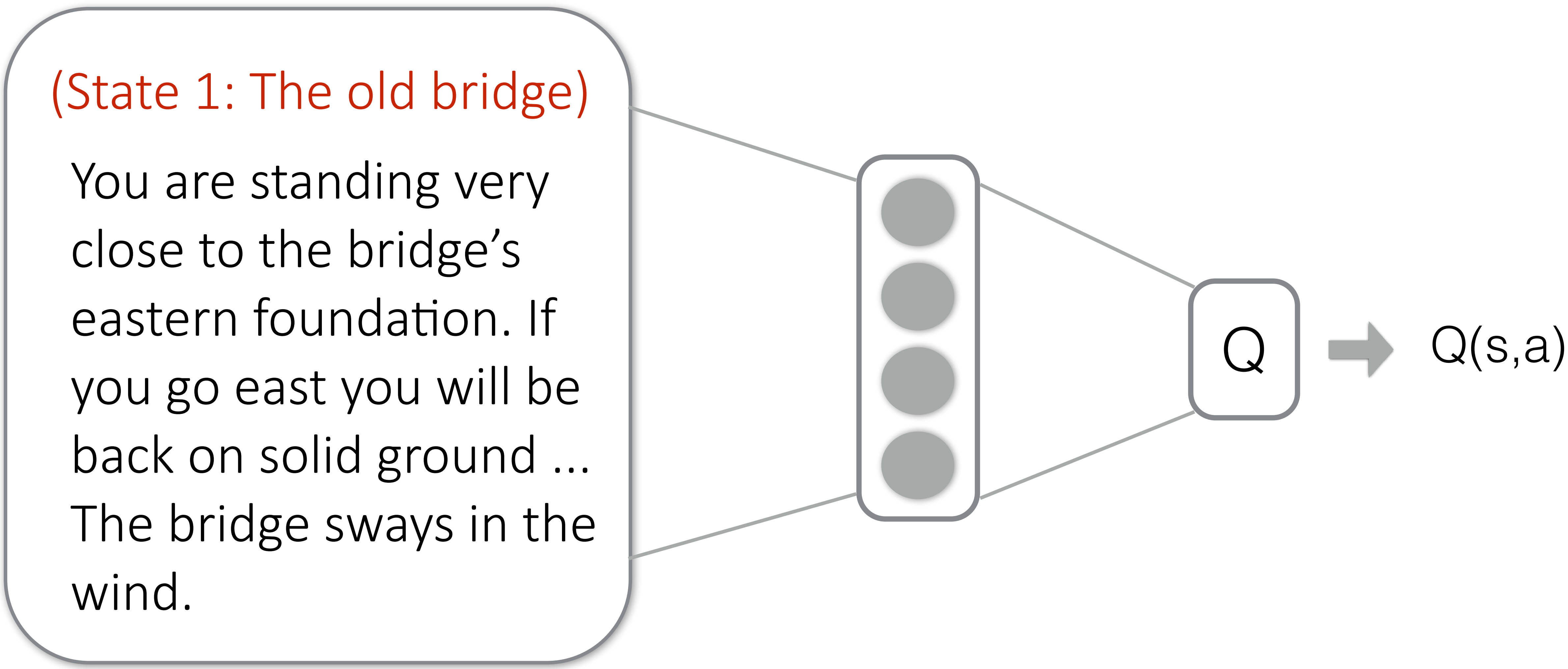

NN for control policy
Input text
T
v
Q
Q values
for all
commands
Recurrent NN to map text to
vector representation

LSTM-DQN



Q(s, a)
Q(s, o)
Linear
Linear
Action-Object Scorer
ReLU
ϕ_A
Linear
v_s
Mean Pooling
Representation Generator
LSTM
LSTM
LSTM
LSTM
ϕ_R
w_1
w_2
w_3
w_n

Algorithm (1)



Obtain Q-values

Algorithm (2)

(State 1: The old bridge)

You are standing very close to the bridge's eastern foundation. If you go east you will be back on solid ground ... The bridge sways in the wind.

$\xrightarrow{a^*}$

Take action using ϵ-greedy

Algorithm (3)

(State 1: The old bridge)

You are standing very close to the bridge's eastern foundation. If you go east you will be back on solid ground ... The bridge sways in the wind.

$\xrightarrow{a^*}$

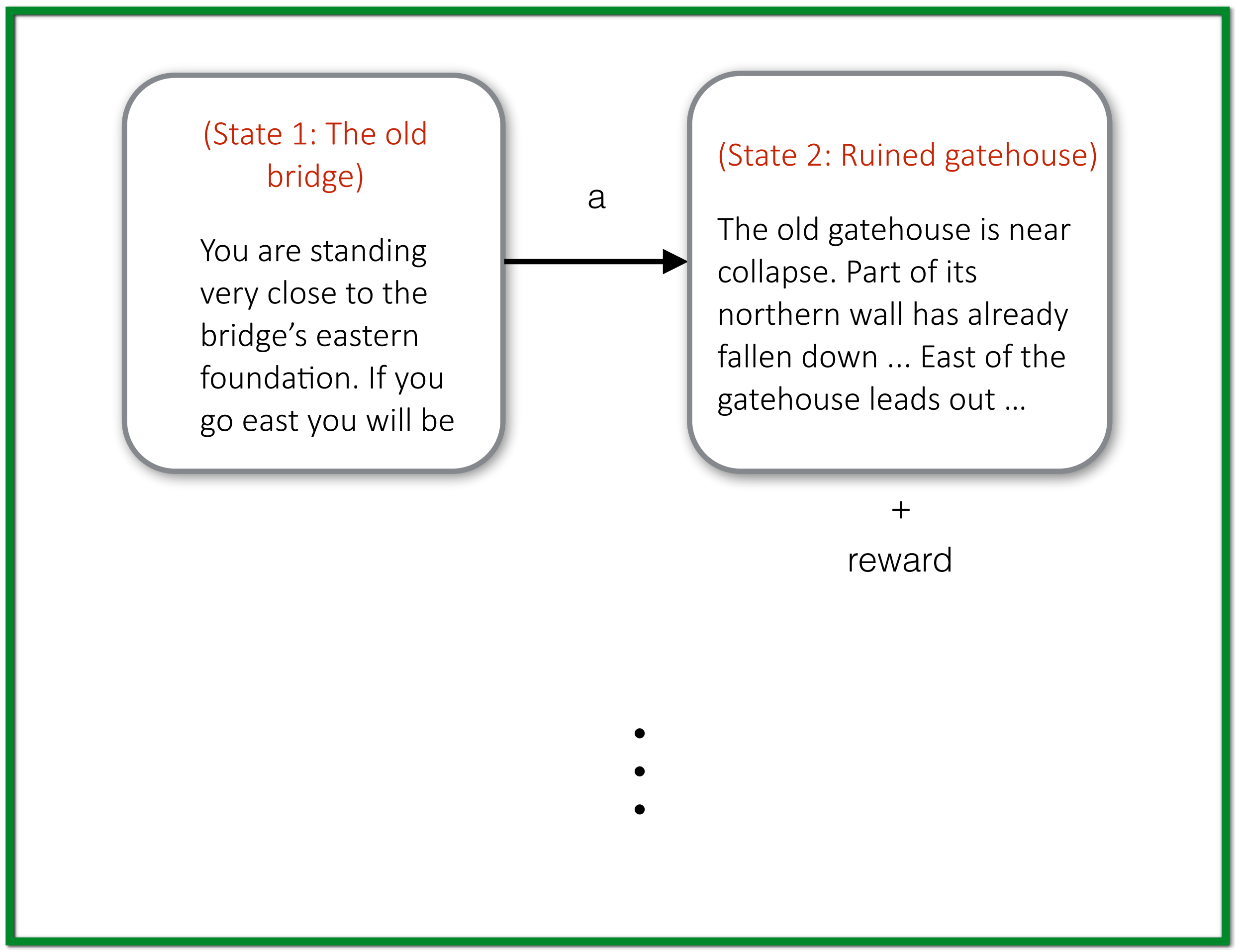
(State 2: Ruined gatehouse)

The old gatehouse is near collapse. Part of its northern wall has already fallen down ... East of the gatehouse leads out ...

\+

reward

Algorithm (4)



~ Sample transitions for updates

Store transition in *experience memory*

Parameter update

(State 1: The old bridge)

You are standing very close to the bridge's eastern foundation. If you go east you will be back on solid ground ... The bridge sways in the wind.

$\xrightarrow{a^*}$

(State 2: Ruined gatehouse)

The old gatehouse is near collapse. Part of its northern wall has already fallen down ... East of the gatehouse leads out ...

+

reward

$$\nabla_{\theta_i} \mathcal{L}_i(\theta_i) = \mathrm{E}_{\hat{s},\hat{a}}[2(y_i - Q(\hat{s}, \hat{a}; \theta_i))\nabla_{\theta_i} Q(\hat{s}, \hat{a}; \theta_i)]$$

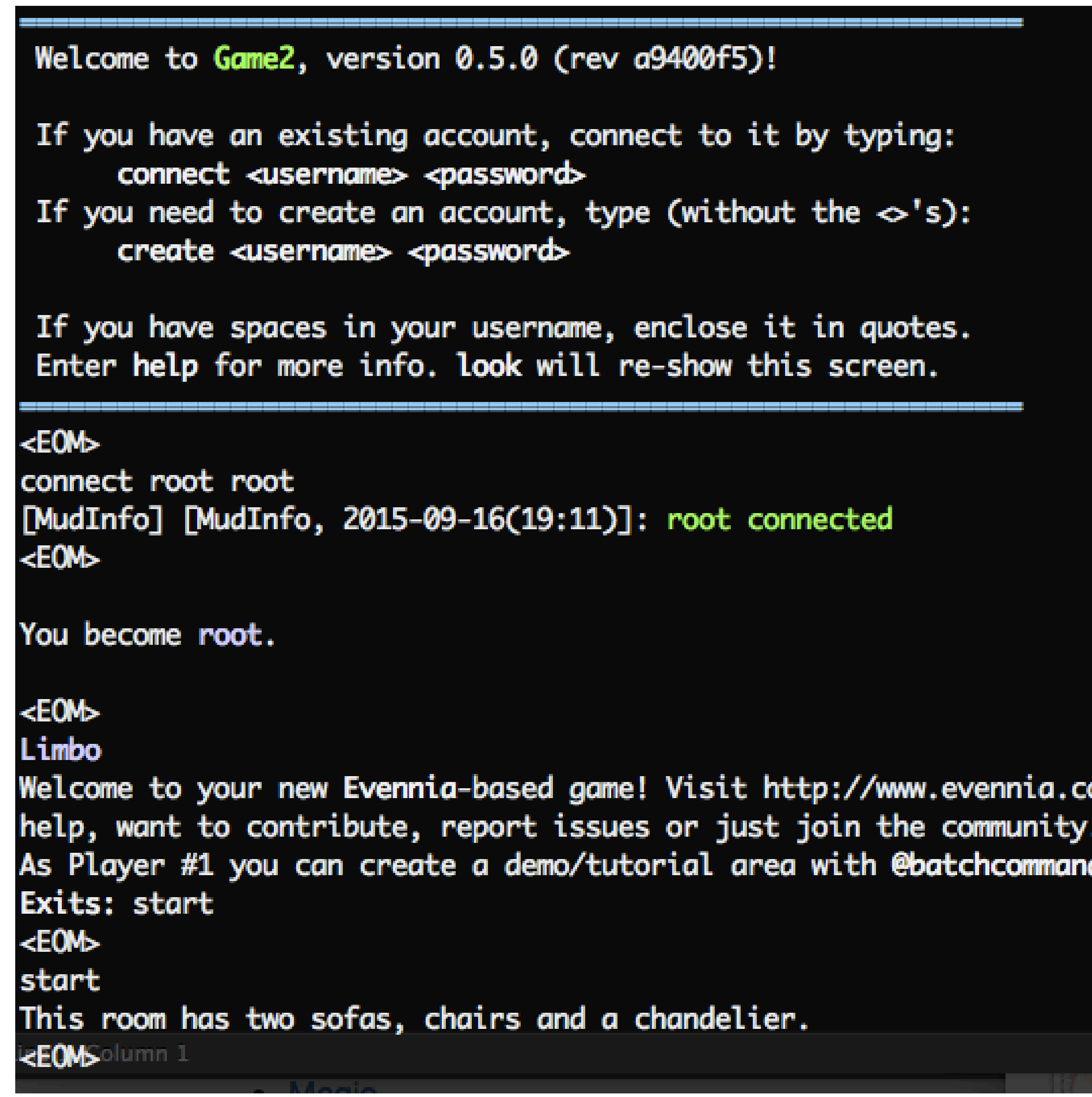
where $$y_i = \mathrm{E}_{\hat{s},\hat{a}}[r + \gamma \max_{a'} Q(s', a'; \theta_{i-1}) \mid \hat{s}, \hat{a}]$$

Game Environment

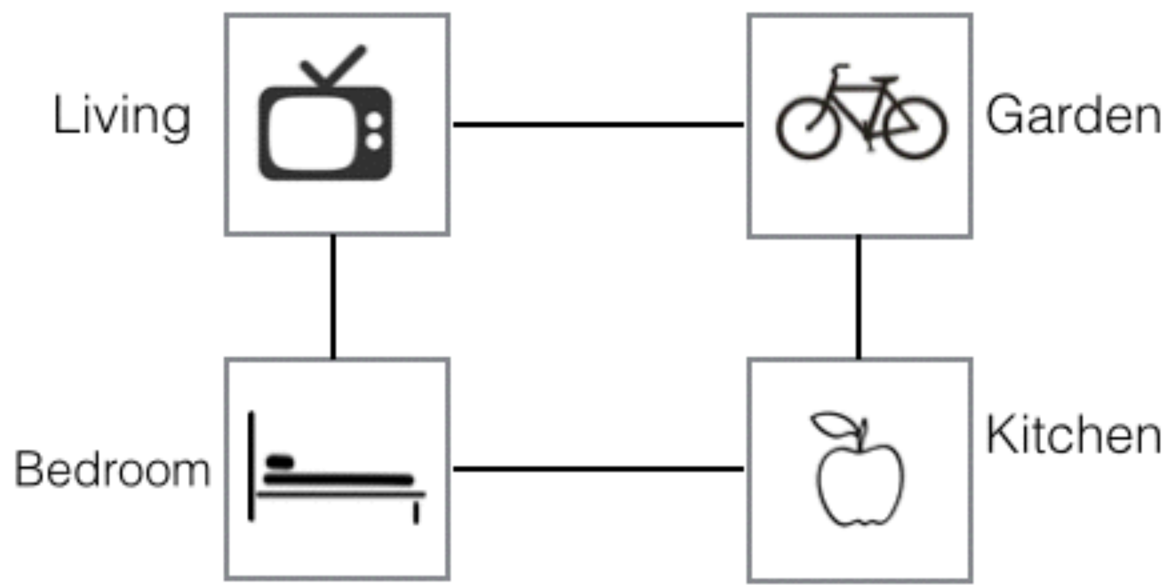
Evennia: a highly extensible python framework for MUD games

Two worlds:

- small game to demonstrate task and analyze learnt representations.
- a pre-existing Fantasy world.

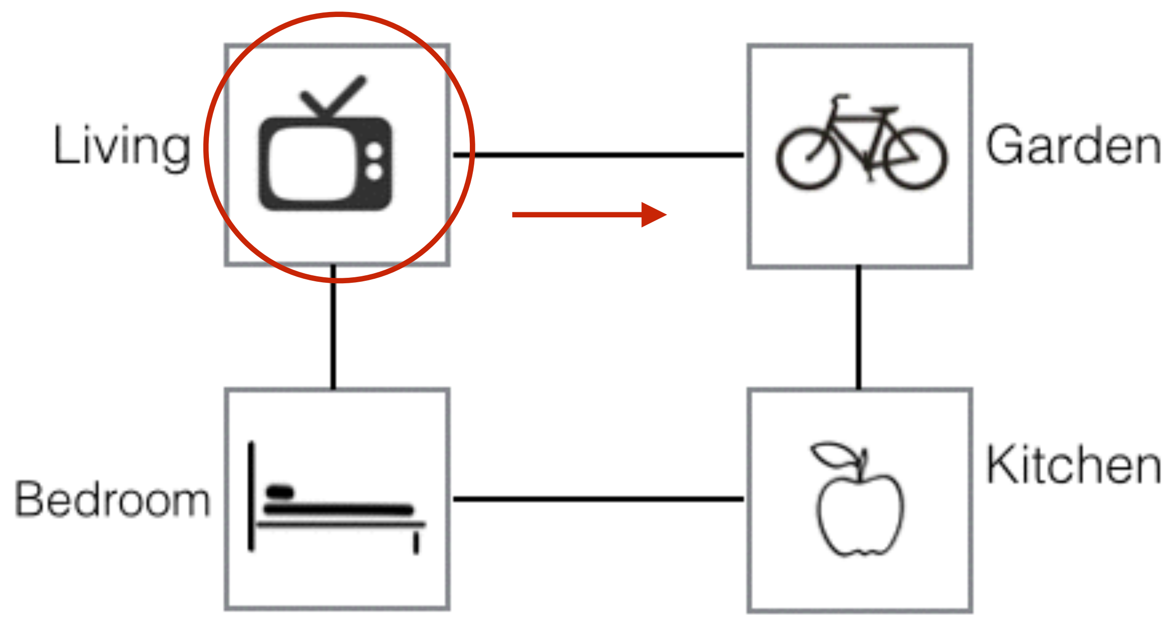


Home World



- Number of different quests: 16
- Vocabulary: 84 words
- Words per description (avg.): 10.5
- Multiple descriptions per room/object.

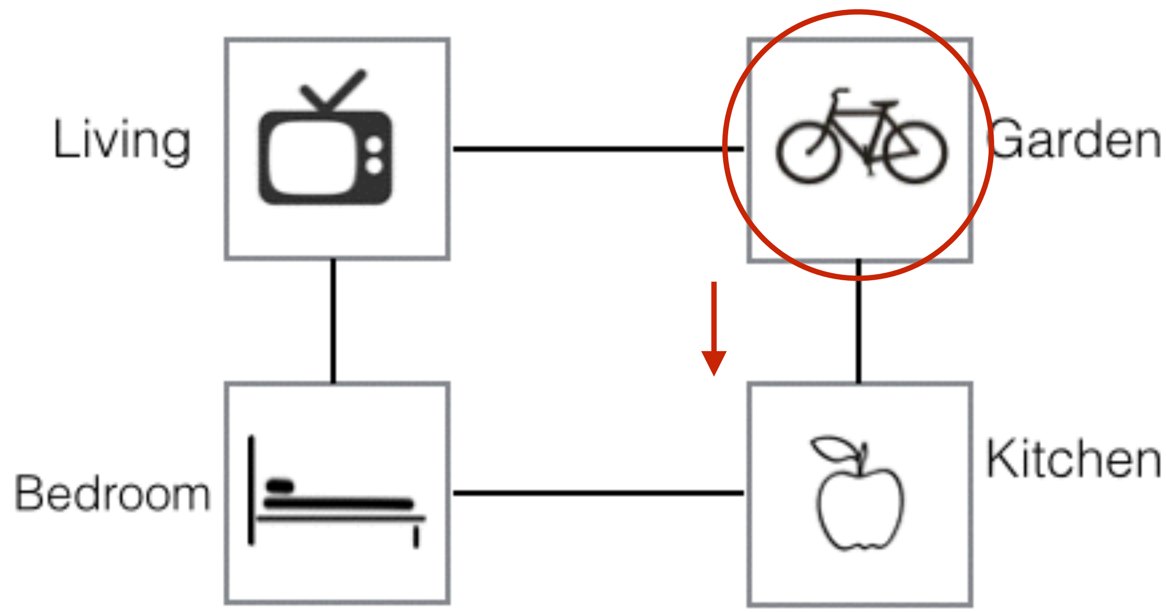
Home World



This room has two sofas, chairs and a chandelier. You are not sleepy now but you are hungry now.

> go east

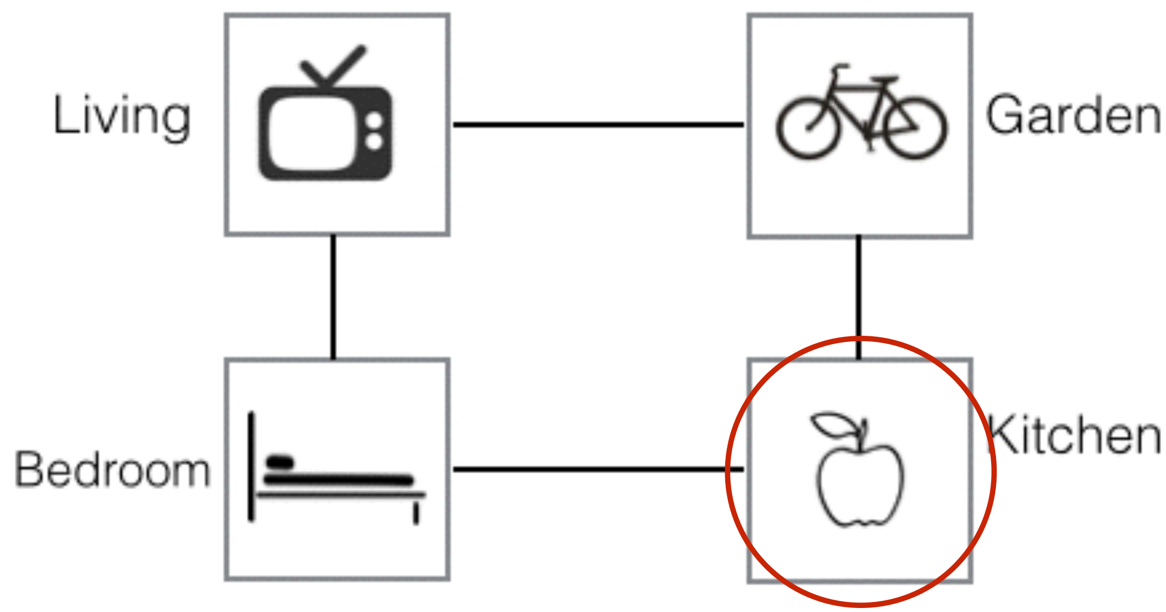
Home World



This area has plants, grass and rabbits. You are not sleepy now but you are hungry now.

> go south

Home World



Reward: +1

You have arrived in the kitchen. You can find food and drinks here. You are not sleepy now but you are hungry now.

> eat apple

Fantasy World

(State 1: The old bridge)

You are standing very close to the bridge's eastern foundation. If you go east you will be back on solid ground ... The bridge sways in the wind.

- Number of rooms: > 56
- Vocabulary: 1340 words
- Avg. no. of words/description: 65.21
- Max descriptions per room: 100

- Considerably more complex
- Varying descriptions per state created by game developers

Evaluation

Two metrics:

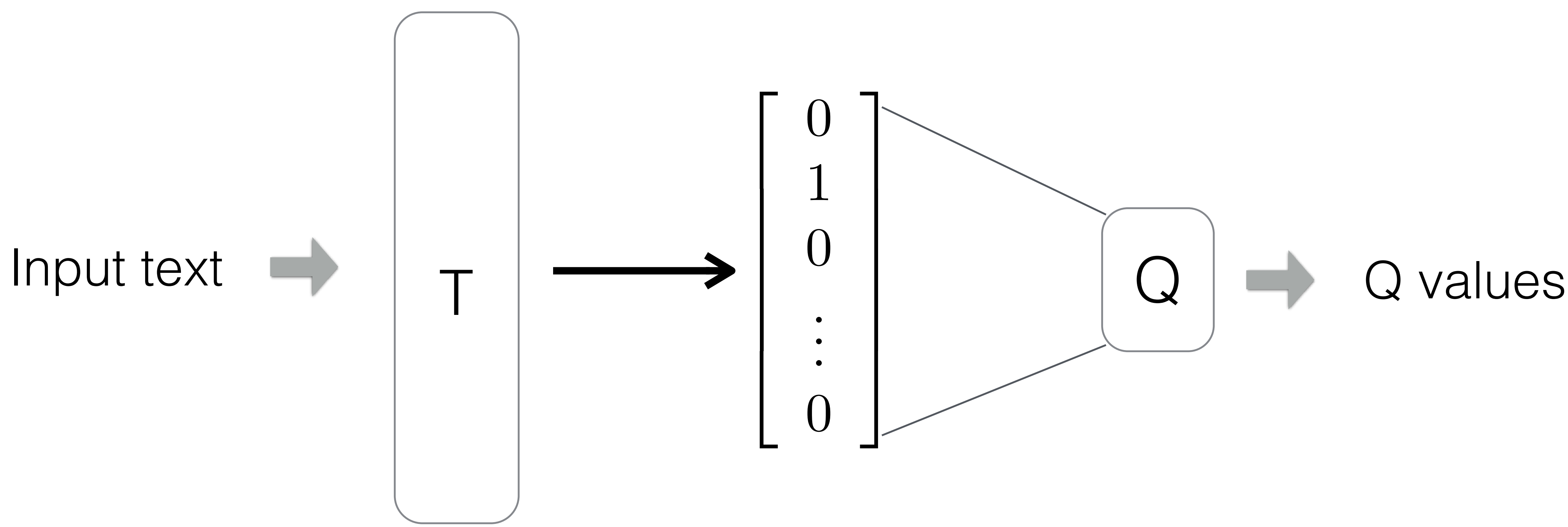
- ✦ Quest completion
- ✦ Cumulative reward per episode
  - Positive rewards for quest fulfillment
  - Negative rewards for *bad* actions

Epoch: Training for *n* episodes followed by evaluation on *n* episodes

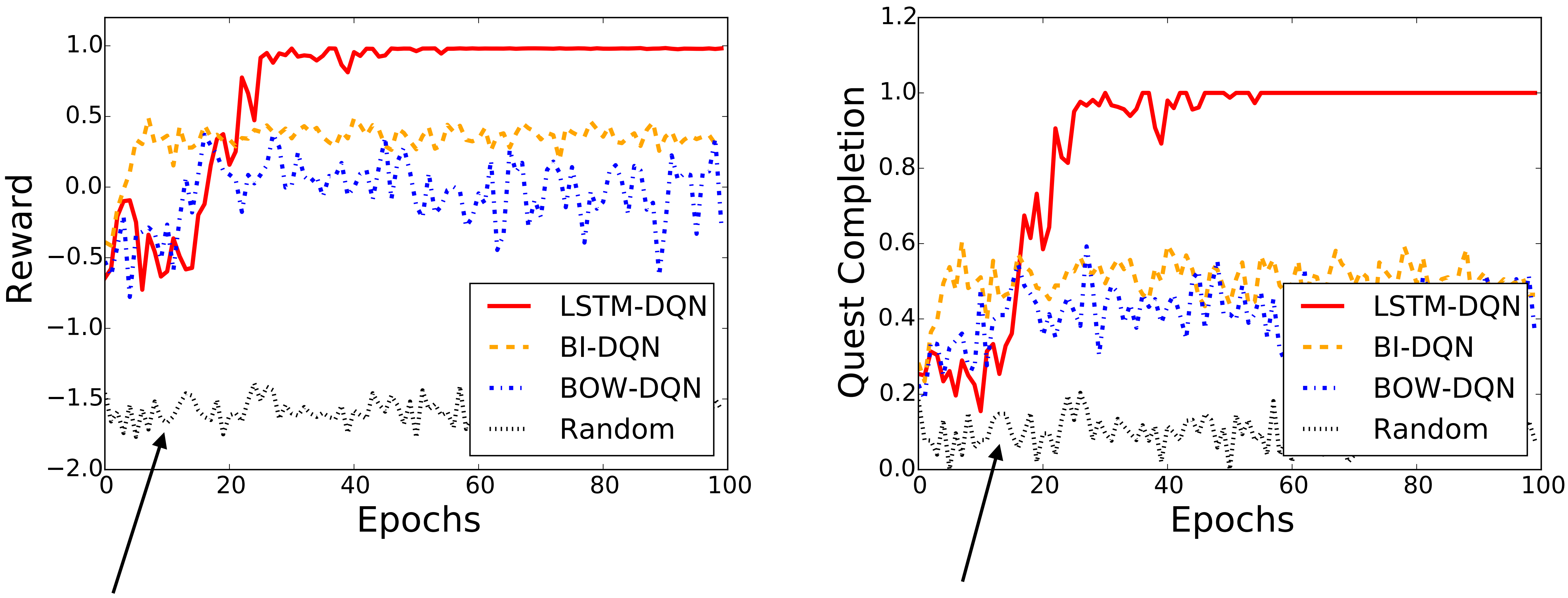
Baselines

- Randomly select actions

- Bag of words: unigrams and bigrams

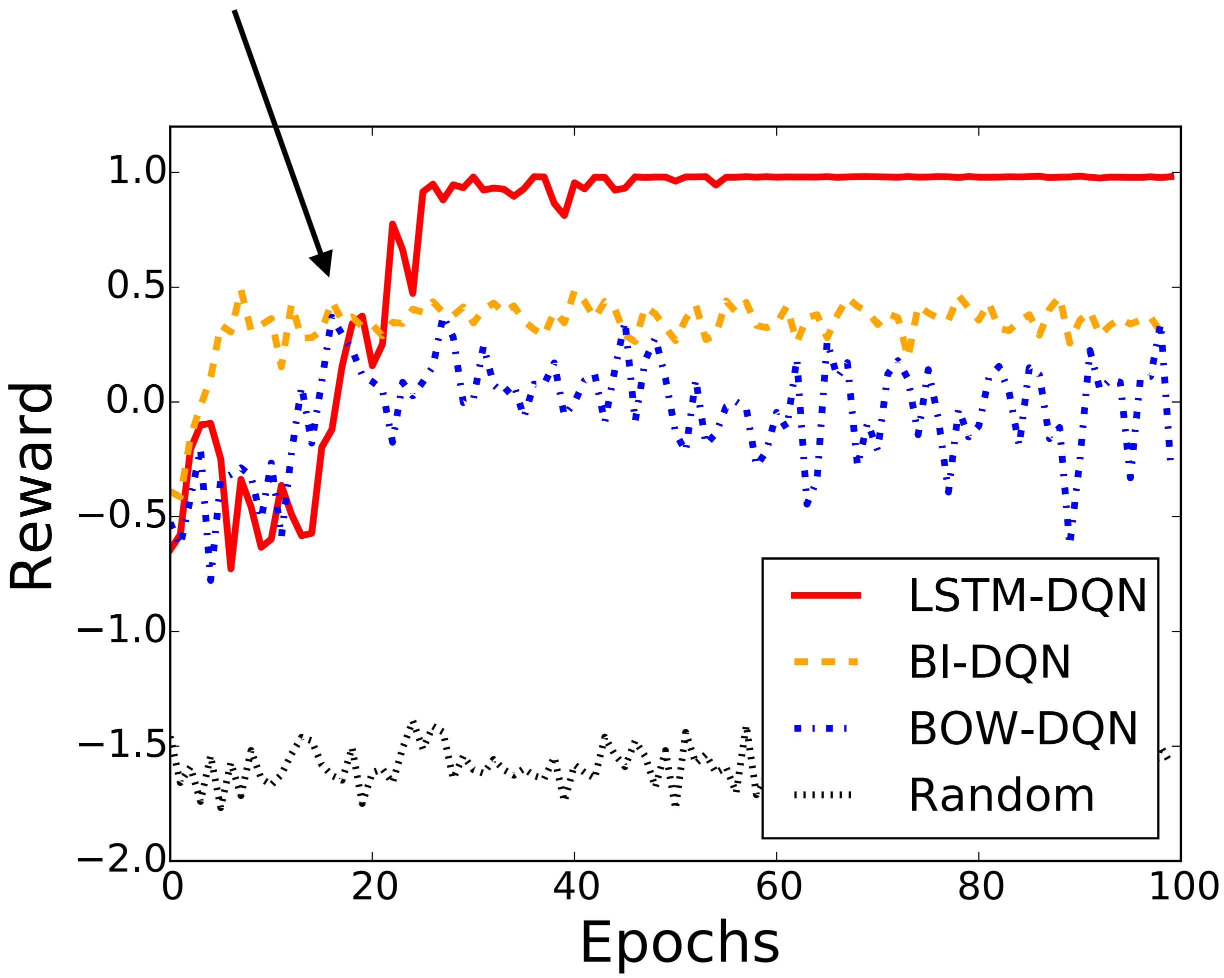


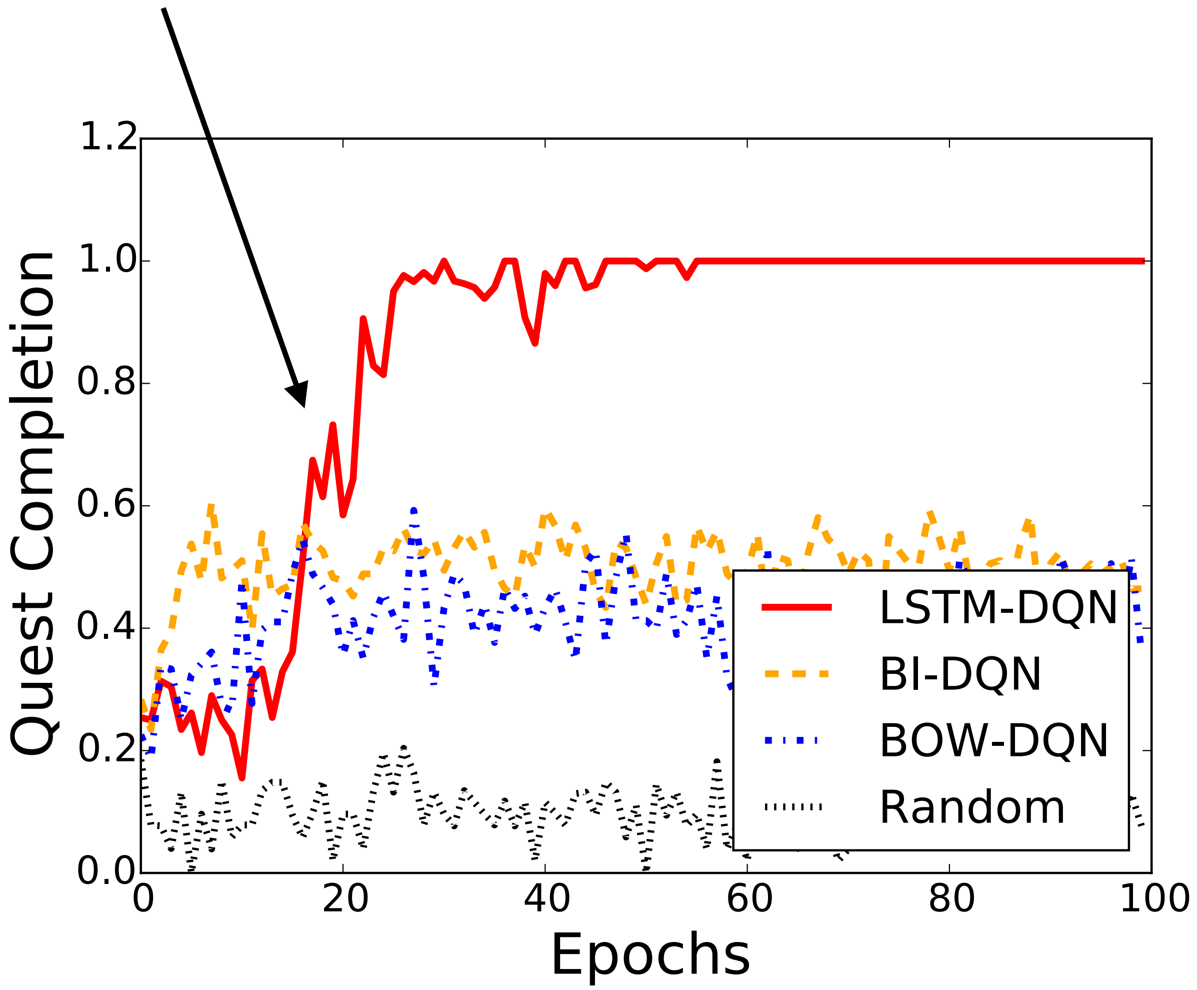
Agent Performance (Home)



Random agent performs poorly

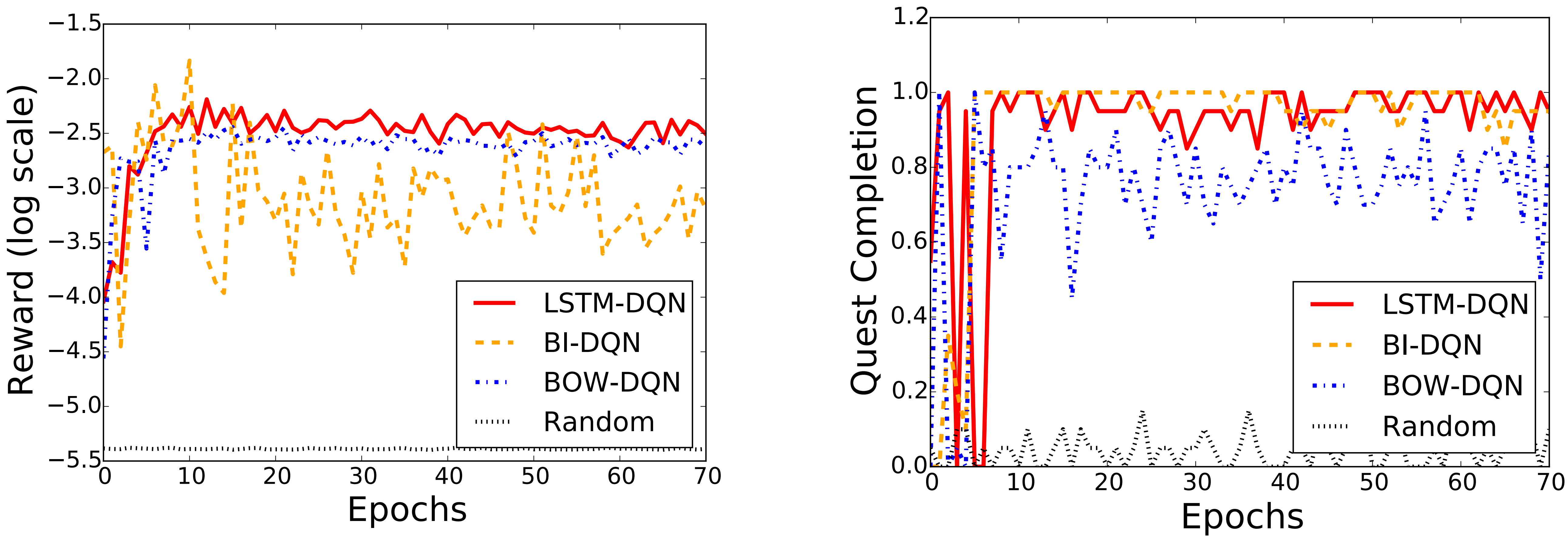
Agent Performance (Home)





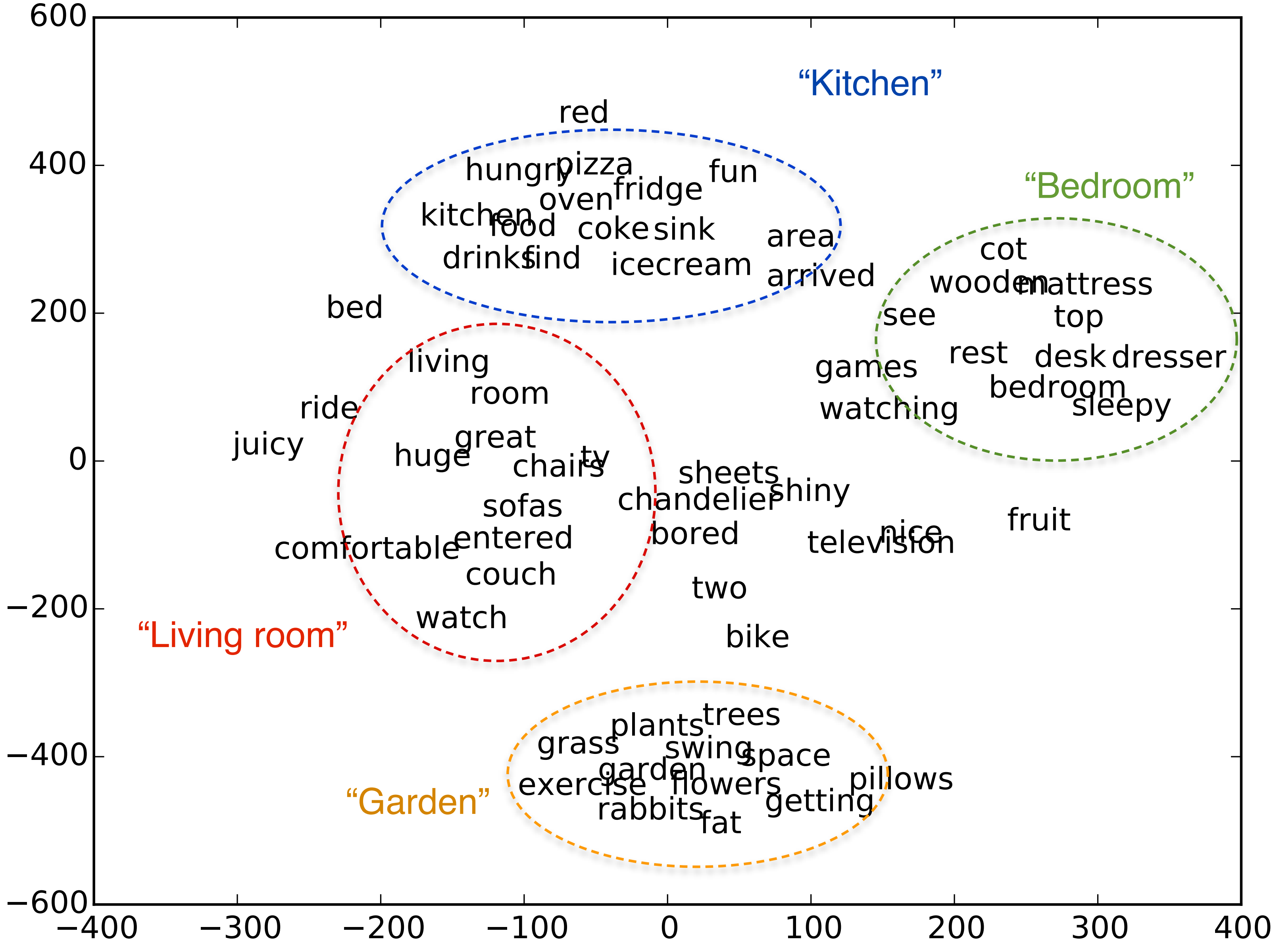
LSTM-DQN has delayed performance jump

Agent Performance (Fantasy)



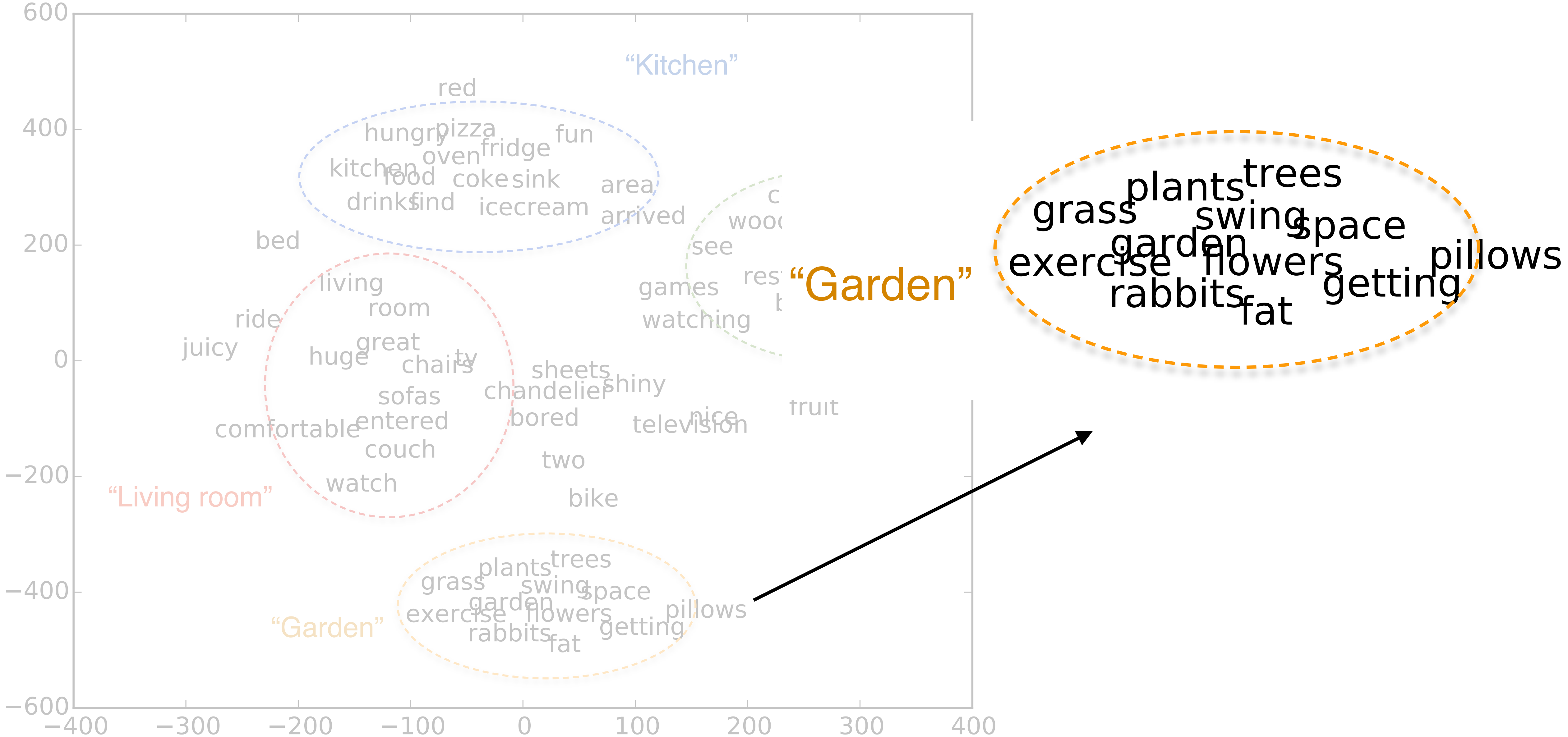
Good representation is essential for successful gameplay

Visualizing Learnt Representations



t-SNE visualization of vectors learnt by agent on Home world

Visualizing Learnt Representations

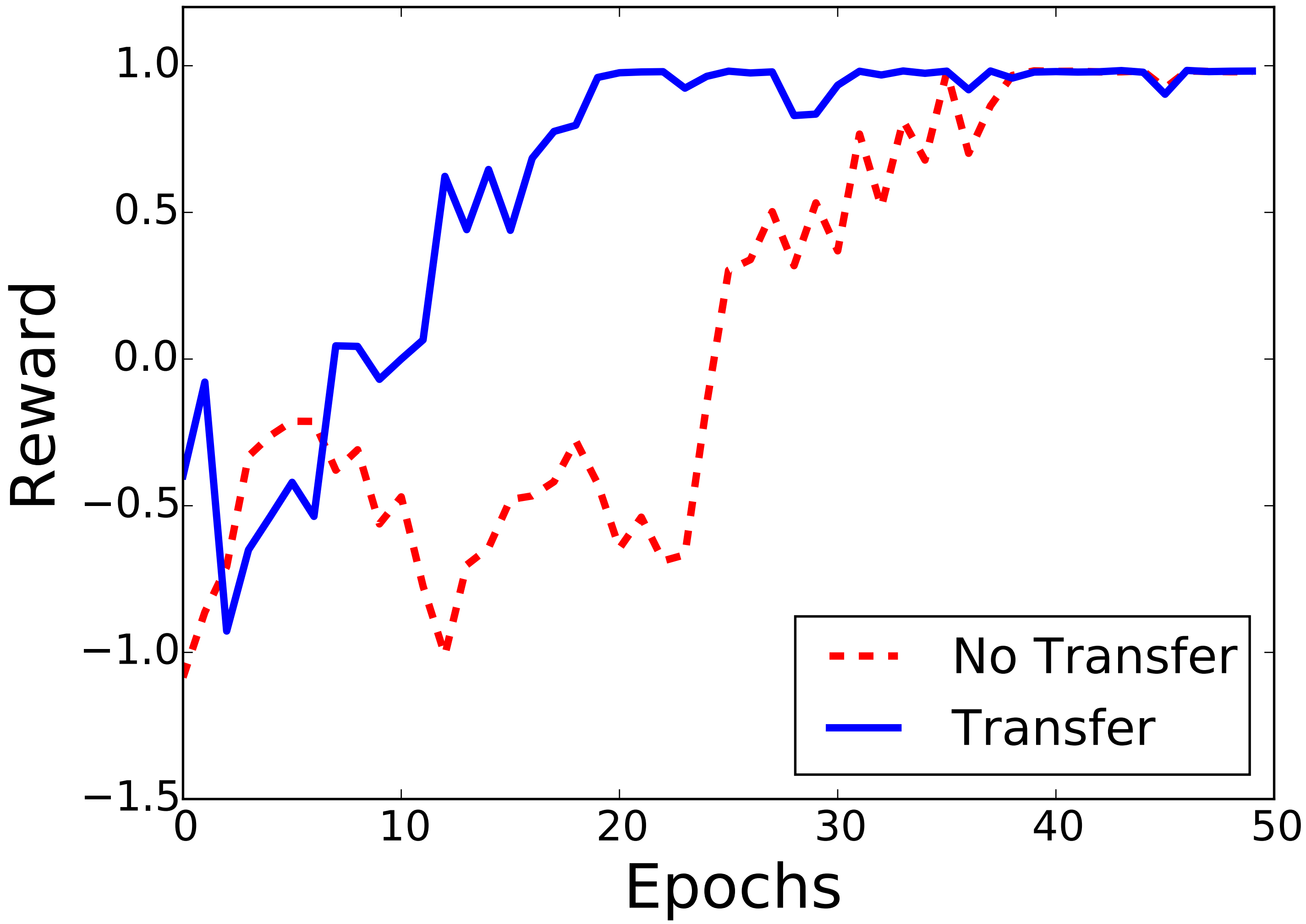


t-SNE visualization of vectors learnt by agent on Home world

Nearby states: Similar representations

Description	**Nearest neighbor**
You are halfways out on the unstable bridge. From the castle you hear a distant howling sound, like that of a large dog or other beast.	The bridge slopes precariously where it extends westwards towards the lowest point - the center point of the hang bridge. You clasp the ropes firmly as the bridge sways and creaks under you.
The ruins opens up to the sky in a small open area, lined by columns. ... To the west is the gatehouse and entrance to the castle, whereas southwards the columns make way for a wide open courtyard.	The old gatehouse is near collapse. .... East the gatehouse leads out to a small open area surrounded by the remains of the castle. There is also a standing archway offering passage to a path along the old southern inner wall.

Transfer Learning (Home)



Play on world with same vocabulary but different physical configuration

Conclusions

- Addressed the task of end-to-end learning of control policies for textual games.
- Learning good representations for text is essential for gameplay.

Code and game framework are available at:
http://people.csail.mit.edu/karthikn/mud-play/