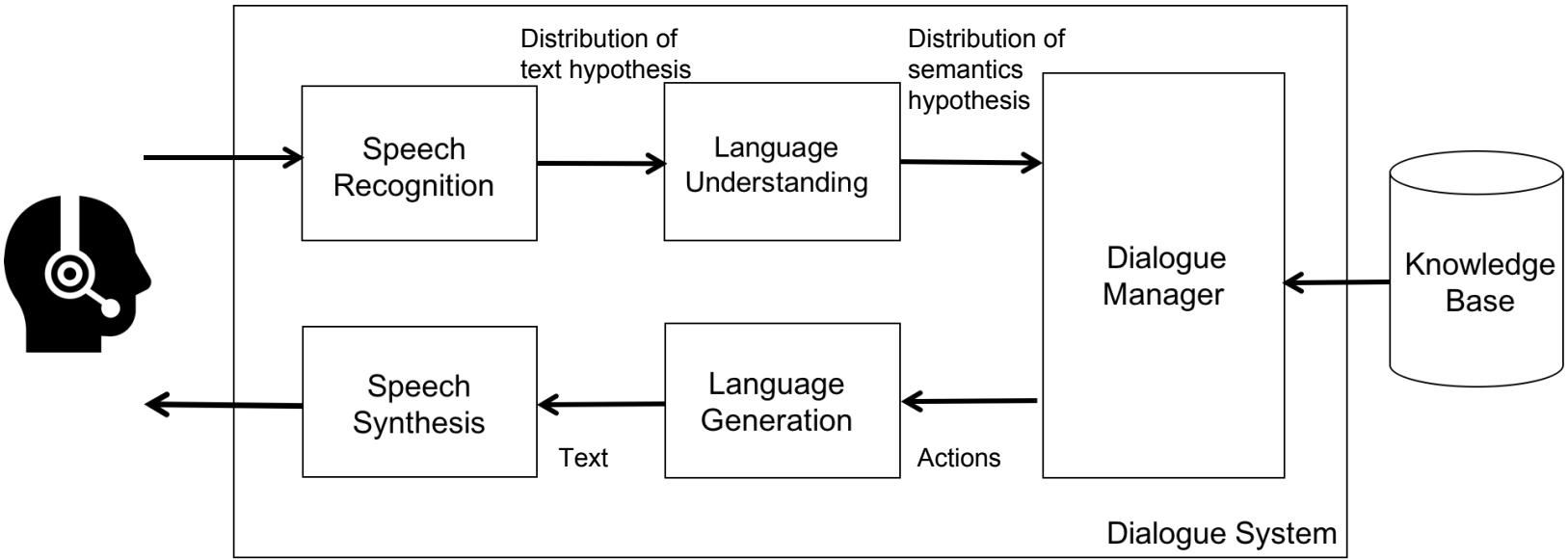
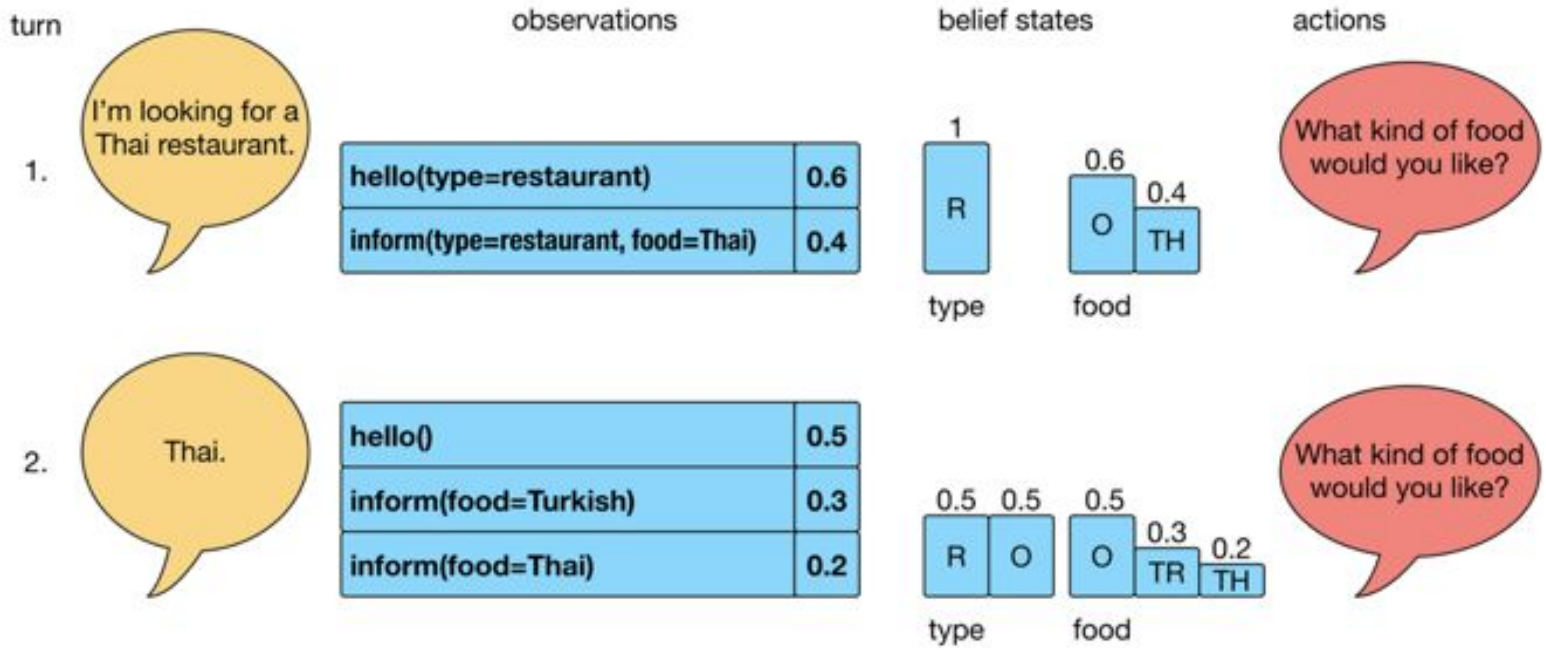
Benchmarking Uncertainty Estimates with Deep Reinforcement Learning for Dialogue Policy Optimisation

Christopher Tegho*, Pawel Budzianowski*, Milica Gasic
Department of Engineering, Cambridge University

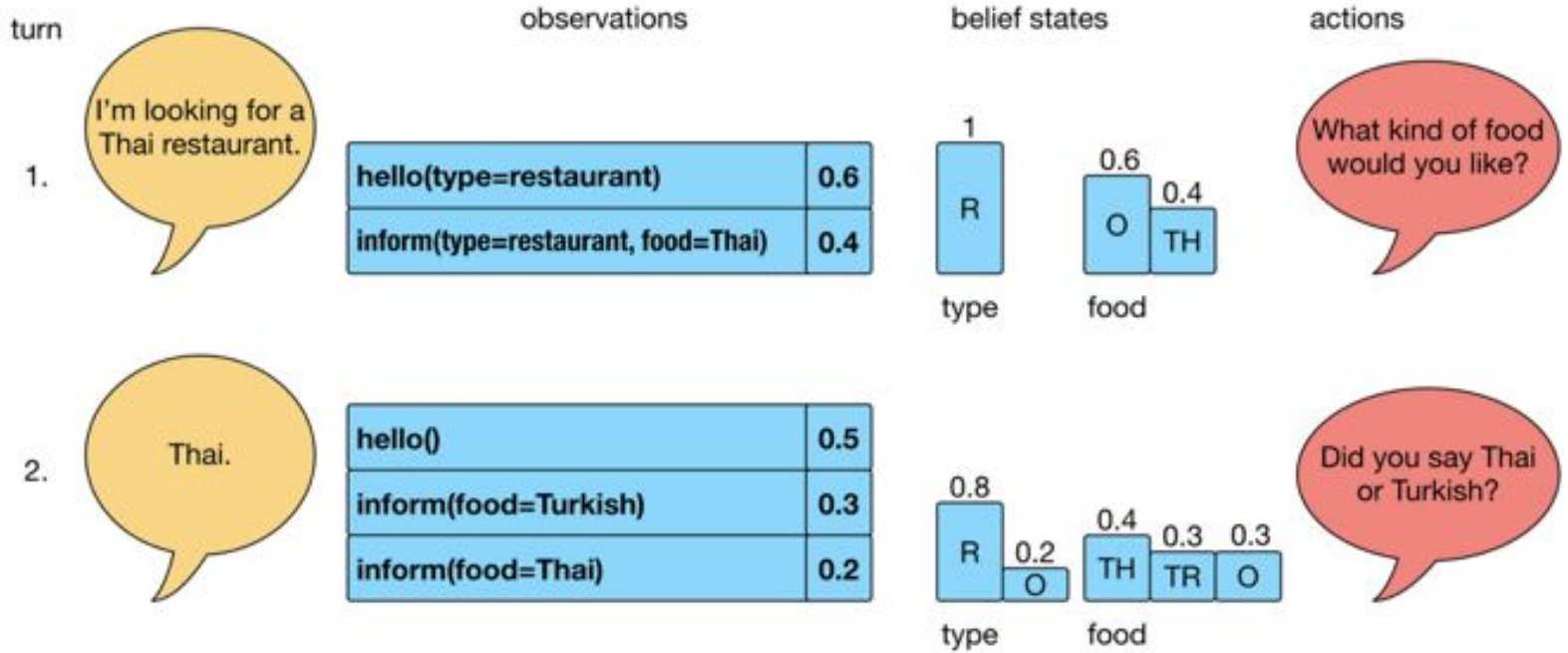
Statistical Dialogue Management Architecture



No Belief State Tracking



Belief State Tracking



Belief State Tracking vs Policy Management

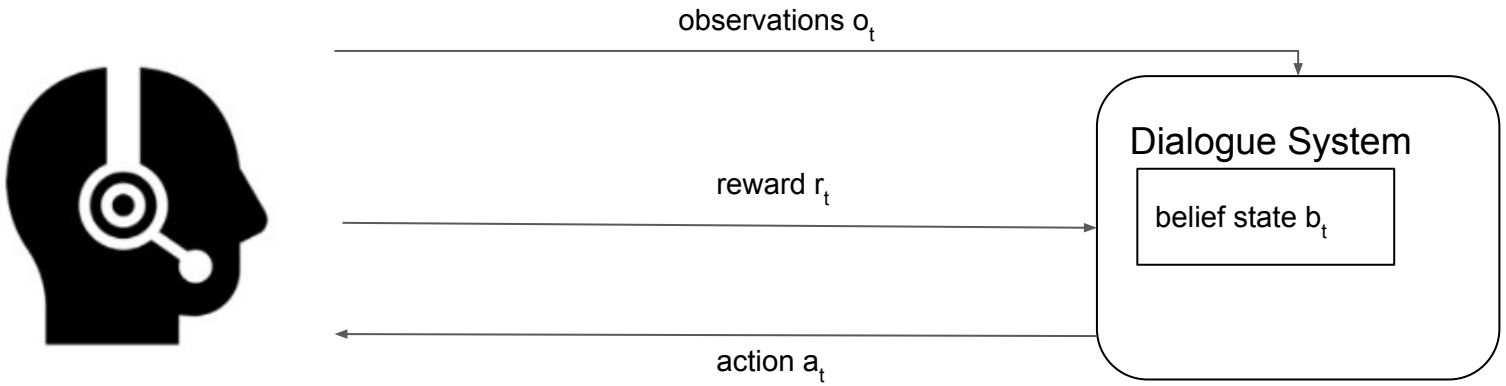
Belief State Tracking

Past: What has happened so far in the dialogue?

Policy Management

Future: What action to take to achieve the best outcome

Reinforcement Learning



Reinforcement Learning

We aim at maximizing a reward obtained along the dialogue:

$$R = \sum_{n=1}^{T} \gamma^t r_t$$

by modelling Q-value function:

$$Q^{\pi}(b, a) = \mathbb{E}_{\pi}\{r_t + \gamma r_{t+1} + \ldots \quad | b_t = b, a_t = a\}$$

Uncertainty Estimates

No random actions
Gaussian distr + figure
Thompson Sampling

- GP SARSA provides an **explicit estimate of uncertainty**, the computational complexity is **cubical**.
- Deep neural network models **scale nicely** with data, but do not provide **an estimate of uncertainty** in vanilla form

Uncertainty Estimates in Neural Networks

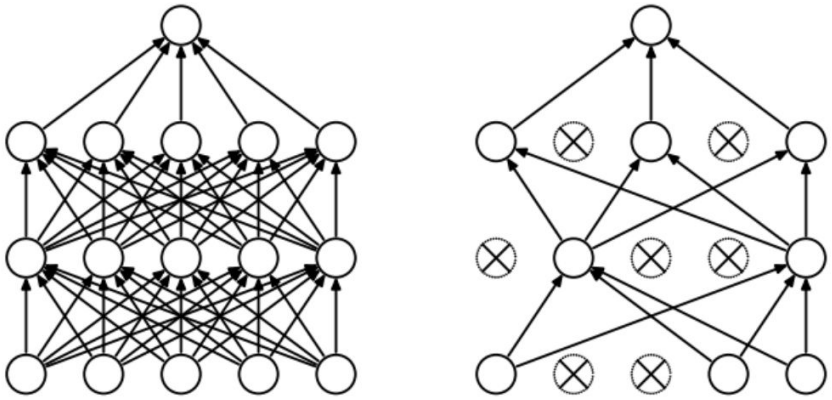
- Problem: Neural networks do not provide **uncertainties** about its estimates.
-
- Number of approached explored - 4 casted in the VI framework

Bayes By Backprop

- All weights are represented by probability distributions over possible values given observed dialogues
- We use sampling-based variational inference. The intractable posterior is approximated with variational posterior.

Uncertainty Estimates in NN

- **Dropout**: Multiply the weight matrix in a given layer by some random noise.
- **Concrete dropout**: Continuously relax the dropout's discrete masks and optimize the dropout probability using gradient methods.



(a) Standard Neural Net (b) After applying dropout.

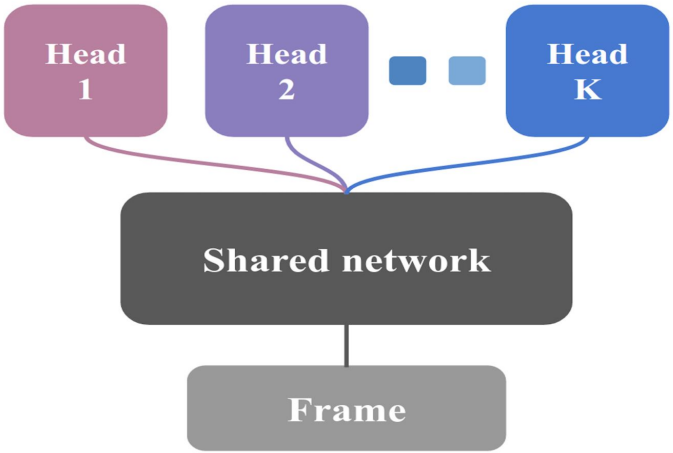
Source: Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting", JMLR 2014

Uncertainty Estimates in NN

- **Alpha-divergence:** The α-divergence measures the similarity between two distributions.

Uncertainty Estimates in NN

Bootstrapped DQN: Several neural networks are randomly initialized which predict in ensemble uncertainty estimates.



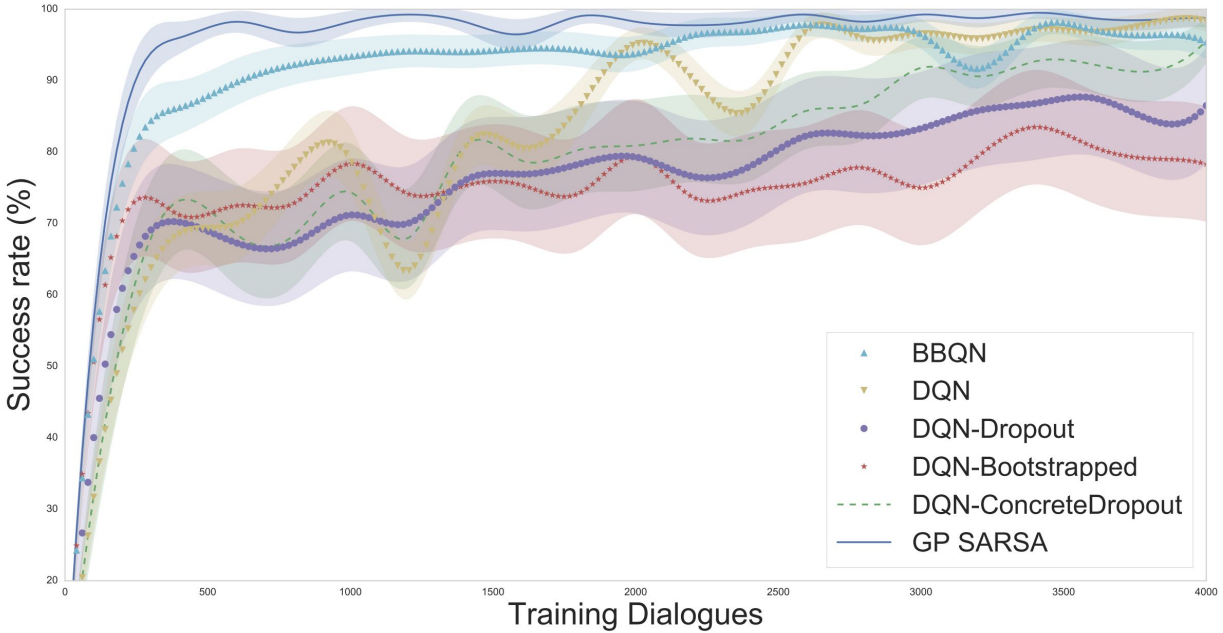
Source: Osband, Ian, Charles Blundell, Alexander Pritzel, and Benjamin Van Roy. "Deep exploration via bootstrapped DQN." In Advances in neural information processing systems, pp. 4026-4034. 2016.

Evaluation Setup

- Cambridge restaurant domain: 100 venues, 6 slots, 3 requestable
- Belief state input of size 268
  (last system act, distribution over user intent …)
- System summary action space of size 14 (inform, request, confirm, …)
- User simulator operating on semantic level, and capable of simulating noise

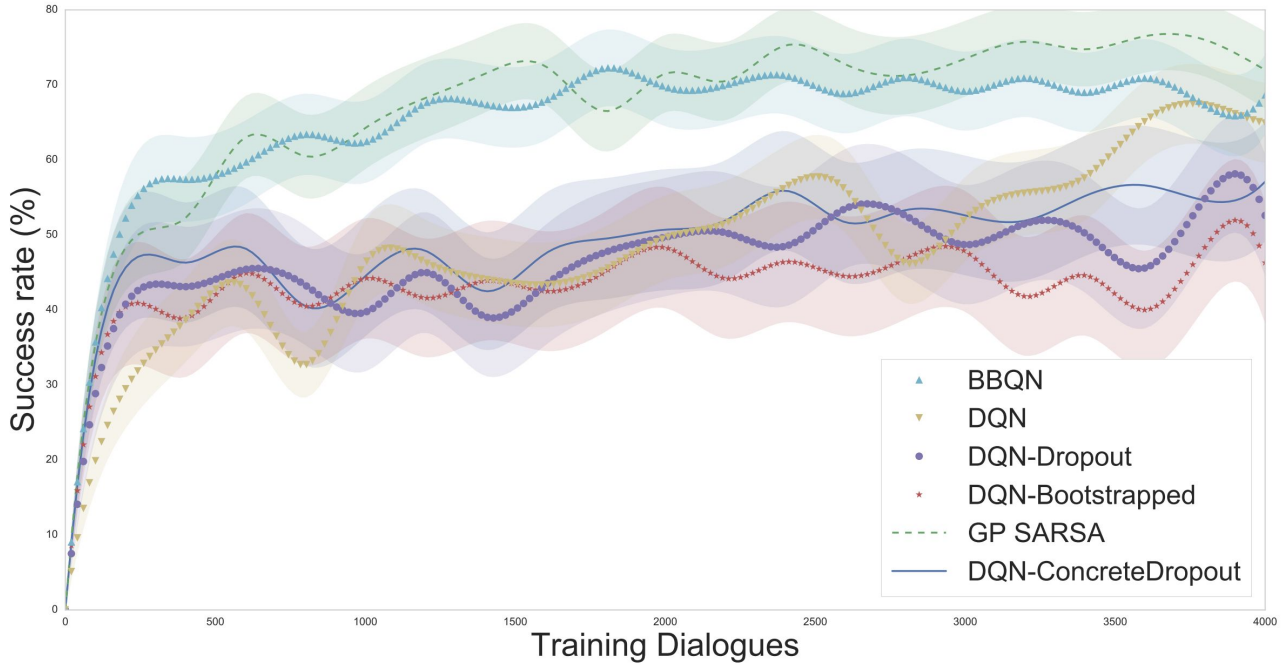
Results

Environment without any noise



Results

Simulated user trained with a 15% semantic error rate, and evaluated on 45% semantic error rate



Conclusion

We train a dialogue agent using reinforcement learning paradigm.

Deep-RL methods proved to be unstable and sample inefficient.

Bayes By Backprop provides uncertainty estimates: more efficient and stable learning is achieved, compared to epsilon-greedy exploration with DQN.

BBQN achieves comparable performance to GPSARSA, especially in more noisy environments, without the cubic computational complexity.