Benchmarking Uncertainty Estimates with Deep Reinforcement Learning for Dialogue Policy Optimisation

Christopher Tegho*, Pawel Budzianowski*, Milica Gasic
Dialogue Systems Group, Department of Engineering, Cambridge University
Statistical Dialogue Management Architecture

Diagram showing the distribution of text hypothesis, semantics hypothesis, speech recognition, language understanding, dialogue manager, knowledge base, speech synthesis, and text generation.
No Belief State Tracking
Reinforcement Learning

observations $o_t$

reward $r_t$

action $a_t$

Dialogue Manager

belief state $b_t$
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:

\[ Q^\pi(b, a) = \mathbb{E}_\pi \{ r_t + \gamma r_{t+1} + \ldots \mid b_t = b, a_t = a \} \]
1. We know how sure the agent is about the action
Uncertainty Estimates

1. We know how sure the agent is about the action
2. We can introduce clever exploration than epsilon-greedy
3. Choosing next action through Thomson Sampling
Uncertainty Estimates

1. We know how sure the agent is about the action
2. We can introduce clever exploration than epsilon-greedy
3. Choosing next action through Thomson Sampling
4. Faster learning, better user experience
Uncertainty Estimates in Neural Networks

- GP SARSA provides an explicit estimate of uncertainty, however, the computational complexity is cubical.
Uncertainty Estimates in Neural Networks

- GP SARSA provides an explicit estimate of uncertainty, however, the computational complexity is cubical.
- Deep neural network models scale nicely with data, but do not provide an explicit estimate of uncertainty.
Uncertainty Estimates in Neural Networks

- GP SARSA provides an **explicit estimate of uncertainty**, however, the computational complexity is **cubical**.
- Deep neural network models **scale nicely** with data, but do not provide an **explicit estimate of uncertainty**
- Uncertainty estimates with NN can be obtained **by approximation**
- Number of approached explored - 4 casted in the **variational inference** 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:
- Loss to minimize:
  $KL(\text{posterior} \mid \text{prior}) - \log \text{likelihood of data}$

Sources:
Uncertainty Estimates in NN

- **Alpha-divergence**: The $\alpha$-divergence measures the similarity between two distributions.
- It’s a generalization over KL divergence
- Bayes By Backprop uses KL divergence, equivalent to Alpha-divergence with alpha = 0

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.

Uncertainty Estimates in NN

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

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
- Capable of simulating noise
Results - Environment without any noise

![Graph showing the success rate of different algorithms over training dialogues. The graph compares various methods like BBQN, DQN, DQN-Dropout, DQN-Bootstrapped, DQN-ConcreteDropout, and GP SARSA.](image)

- BBQN
- DQN
- DQN-Dropout
- DQN-Bootstrapped
- DQN-ConcreteDropout
- GP SARSA
Results - Environment with noise
Conclusion

We train a dialogue agent using reinforcement learning paradigm.

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

We train a dialogue agent using reinforcement learning paradigm.

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

We tested 5 different approaches to introduce uncertainty estimates into Deep-RL agent.

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