



Back-off Action Selection in Summary Space-Based POMDP Dialogue Systems

M. Gašić, F. Lefèvre, F. Jurčićek, S. Keizer, F. Mairesse, B. Thomson, K. Yu, S. Young

Cambridge University Engineering Department | {mg436, frf12, fj228, sk561, farm2, brmt2, ky219, sjy}@eng.cam.ac.uk

Introduction

Partially Observable Markov Decision Process (POMDP) approach

- enables modelling the uncertainty that arises from the speech understanding errors

To achieve **tractability**

- both state and action spaces must be reduced to smaller scale summary spaces where learning is tractable

Mapping back from summary space

- can lead to an invalid action in the master space

The invalid action problem

The invalid action problem in discrete MDP

- State is known
- For each state a subset of valid actions can be considered during learning

The invalid action problem in POMDP

- State is unknown
- Belief state (distribution over states) is maintained
- Any state is possible, therefore a subset of valid actions cannot be defined for each belief state

Example

System	Summary Action	HELLO
	Master Action	hello() <i>Hello, how may I help you?</i>
User		<i>Um, something cheap...</i>
	Recognition	IS IT CHEAP
	Semantics	confirm(pricerange=cheap)
System	Summary Action	INFORM
	Master Action	invalid nothing to inform about back off to REPEAT repeat() <i>Can you please repeat that?</i>

Is there a better back-off action?

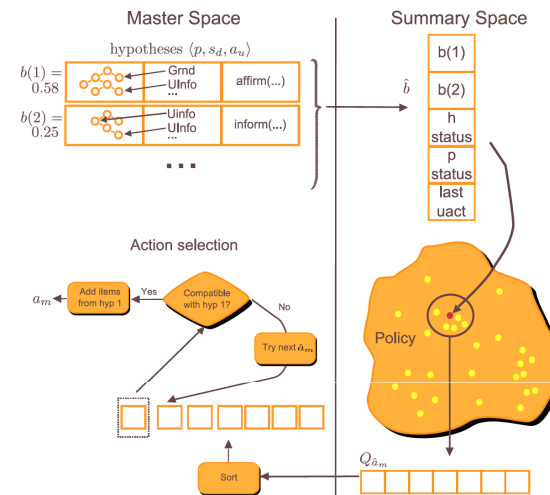
N-best Back-off Selection

- If the validity of proposed action can be assessed during execution, a back-off action can be defined
- Reinforcement learning algorithm can be adapted to order actions in an N-best list of back-off actions
- Thus guaranteeing that the action that is eventually performed gives the highest expected reward

HIS Dialogue Manager

Hidden Information State (HIS) dialogue system is a trainable and scalable implementation of a dialogue system based on POMDP approach.

- Represents goals as branching trees
- Partitions the state space and retains belief over partitions
- Incorporates user and observation model in belief updates
- Maintains two state spaces:
 - Updates belief in master space
 - Optimises policy in summary space using grid-based approach



Extended MCC algorithm

To train the policy producing the optimal action list for each summary state, a grid-based Monte Carlo Control (MCC) algorithm is used.

- A summary point is approximated with its closest grid point
- Q-values are estimated for each grid point
- Exploitation: perform the first valid actions from the list of actions ordered by Q-values
- Exploration: perform the first valid action from a list of randomly ordered actions
- The reward is assigned to the action that was actually taken

Experiments

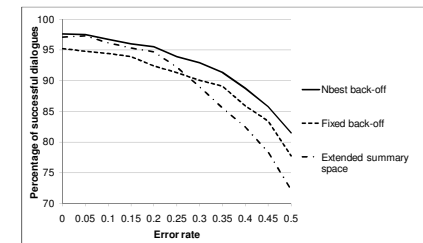
Several strategies were compared:

Fixed back-off – one action that can always be performed is used for back-off (repeat)

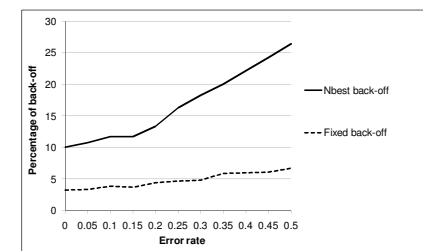
Extended Summary Space – the summary space is extended with validity information for each action, no back-off is needed

N-best back-off – extended MCC is used to propose the ordered list of back-off actions

N-best back-off outperforms other strategies:



Backing off contributes more to robustness in the N-best back-off strategy than in the Fixed back-off strategy:



Conclusion

The N-best back-off selection approach provides an effective solution to the invalid action problem without expanding the summary space.

Acknowledgements

This research was partly funded by the UK EPSRC under grant agreement EP/F013930/1 and by the EU FP7 Programme under grant agreement 216594 (CLASSIC project: www.classic-project.org).

