Machine Learning of Effective Dialogue Management Policies

Oliver Lemon, James Henderson, Kallirroi Georgila. HCRC/ICCS, School of Informatics

`{jhender6,olemon,kgeorgil}@inf.ed.ac.uk`

`http://www.talk-project.org`

Abstract

We are investigating machine learning methods for robust and effective dialogue management policies. In the **TALK** project (an EU FP6 project), we have developed a novel combination of reinforcement learning and supervised learning, which allows us to learn an entire dialogue policy from a fixed corpus of human-machine dialogues. We have also developed *user simulations* for use in automatic evaluation and optimization of policies. Experiments with human users have demonstrated the advantages of the learned policy over state-of-the-art hand-coded policies (+14.4% average reward). In a new EPSRC project, "***End-to-end Integrated Statistical Processing for Context-Aware Dialogue Systems***", we will extend this work by developing tractable and effective techniques for the integrated treatment of uncertainty in context-aware dialogue systems, for example using Partially Observable MDPs.

Motivation

Dialogue systems are gaining increased commercial and social importance, but current dialogue systems use hand-coded dialogue management policies, with the resulting development costs and lack of robustness. One promising approach is to use statistical machine learning, but previous work on dialogue has not identified tractable and effective methods for learning complex dialogue strategies with large state spaces.

The TALK Project

The EU FP6 project, "***TALK: Talk and Look, Tools for Ambient Linguistic Knowledge***":

- Reinforcement Learning for dialogue management
- Context-sensitive speech recognition
- Multilingual and multimodal development tools
- Reconfigurable dialogue systems using ontologies

Learning Strategies

Using reinforcement learning to learn dialogue management policies: [3]:

- Previous work has focused on small state spaces and small sets of actions.
- We address learning to choose between a relatively large number of dialogue actions (70), with a very large state space (over 10^{87} states are theoretically possible).
- We use linear function approximation to handle the large state space.
- We propose a "hybrid learning" method to search through the huge space of possible policies despite having only a limited corpus of dialogues (697 dialogues in our experiments).

COMMUNICATOR Corpus

We extended the annotation of the 2001 COMMUNICATOR corpus of human-machine dialogues. We *automatically* annotated this corpus with a very **rich representation of dialogue state** [2]. These complex representations pose computational problems for traditional reinforcement learning methods, but by overcoming these problems we have gained a very powerful paradigm for learning dialogue strategies.

```
Speaker: user
Asr Input: october three first late morning
Speech Act: [provide_info]
Task: [depart_time]
Filled Slot: [depart_time]
Filled Slot Value: [late morning]
Confirmed Slot: [destin_city]
Filled Slot History: [origin_city], [destin_city], [depart_time]
Confirmed Slot History: [ ], [origin_city], [destin_city]
```

Evaluations

We first tested our "hybrid learning" model using **simulated dialogues** [1]:

- Trained the hybrid model on annotated dialogues from the COMMUNICATOR corpus.
- Trained user simulations on the user actions of the annotated COMMUNICATOR corpus, using supervised learning.
- Ran the hybrid system model against the user simulations to produce simulated dialogues.
- Scored the simulated dialogues using a combination of task completion and dialogue length.

System	total score	filled slots	confirmed slots	length penalty
hybrid RL/SL	**140.3**	88.0	70.0	-17.7
pure SL	**138.3**	89.2	69.1	-20.0
all COMMUNICATOR	**127.1**	84.5	63.9	-21.3
pure RL	**34.9**	56.9	8.3	-31.3

Figure 1: Average scores after the first flight offer

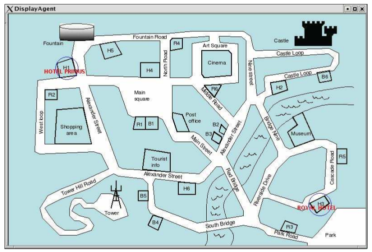


Figure 2: The TALK TownInfo System

We then tested the hybrid learning model with **real users** [4]:

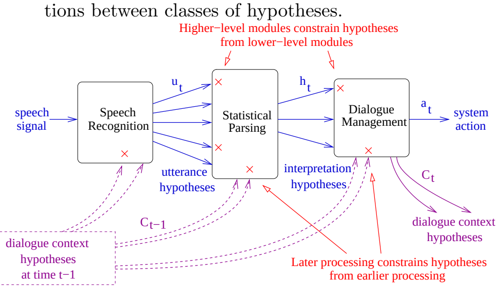
- Ported the trained hybrid model to the tourist information domain (figure 2).
- Compared against a hand-coded baseline system, and another learned model developed at Cambridge University.
- Ran controlled experiments with human users.

Policy	Percieved task completion	User pref.	System turns	Reward
hybrid RL/SL	81.8%	2.67	11.6	**74.9**
hand-coded	67.6%	2.75	14.9	**60.5**

End-to-end Modeling of Uncertainty

A new **EPSRC** project, starting January 2007, "***End-to-end Integrated Statistical Processing for Context-Aware Dialogue Systems***":

- Uncertainty pervades **every module** of a dialogue system, resulting in uncertainty about the **state of the dialogue**.
- Pursuing every possible hypothesis is **not tractable**.
- We will combine:
  - the compact representations of uncertainty developed within approaches to dialogue managment
  - with n-best lists to allow for arbitrary disjunctions between classes of hypotheses.



References

[1] Kallirroi Georgila, James Henderson, and Oliver Lemon. User simulation for spoken dialogue systems: Learning and evaluation. In *Proceedings of Interspeech/ICSLP 2006*, 2006.

[2] Kallirroi Georgila, Oliver Lemon, and James Henderson. Automatic annotation of COMMUNICATOR dialogue data for learning dialogue strategies and user simulations. In *Ninth Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL: DIALOR)*, 2005.

[3] James Henderson, Oliver Lemon, and Kallirroi Georgila. Hybrid Reinforcement/Supervised Learning for Dialogue Policies from COMMUNICATOR data. In *IJCAI workshop on Knowledge and Reasoning in Practical Dialogue Systems*, 2005.

[4] Oliver Lemon, Kallirroi Georgila, and James Henderson. Evaluating Effectiveness and Portability of Reinforcement Learned Dialogue Strategies with real users: the TALK TownInfo Evaluation. In *Spoken Language Technology*, page (to appear), 2006.

School of informatics