Machine Learning for Spoken Dialogue Systems

Oliver Lemon

University of Edinburgh
School of Informatics
2 Buccleuch Place, Edinburgh, EH8 9LW - UK
olemon@inf.ed.ac.uk

Olivier Pietquin

École Supérieure d'Électricité
Metz Campus - IMS Research Group
2 rue Edouard Belin, F-57070 Metz - France
olivier.pietquin@supelec.fr

Abstract

During the last decade, research in the field of Spoken Dialogue Systems (SDS) has experienced increasing growth. However, the design and optimization of SDS is not only about combining speech and language processing systems such as Automatic Speech Recognition (ASR), parsers, Natural Language Generation (NLG), and Text-to-Speech (TTS) synthesis systems. It also requires the development of dialogue strategies taking at least into account the performances of these subsystems (and others), the nature of the task (e.g. form filling, tutoring, robot control, or database search/browsing), and the user's behaviour (e.g. cooperativeness, expertise). Due to the great variability of these factors, reuse of previous hand-crafted designs is also made very difficult. For these reasons, statistical machine learning (ML) methods applied to automatic SDS optimization have been a leading research area for the last few years. In this paper, we provide a short review of the field and of recent advances.

Index Terms: Spoken Dialogue Systems, Machine Learning

1. The promise of ML techniques

Levin, Pieraccini, and Eckert [1] were the first researchers to model human-machine dialogue as a Markov Decision Process (MDP), thus making development of spoken dialogue systems amenable to machine learning approaches. Ten years later, automatic learning of optimal dialogue strategies is now a leading domain of research, with several recent advances being driven by EC funding (see e.g. the TALK project www.talk-project.org). Among machine learning techniques for spoken dialogue strategy optimization, reinforcement learning [2] using Markov Decision Processes (MDPs) [1, 3, 4, 5, 6, 7] and Partially Observable MDP (POMDPs) [8, 9, 10] has become a particular focus.

Why is this? Statistical computational learning approaches offer several key potential advantages over the standard rule-based hand-coding approach to dialogue systems development:

- data-driven development cycle
- provably optimal action policies
- a precise mathematical model for action selection
- possibilities for generalization to unseen states
- reduced development and deployment costs for industry.

However, it is worth noting that several aspects of the MDP approach have been questioned, e.g. [11].

2. Problems and approaches

2.1. Tractability, and dimensionality reduction methods

One of the most pressing issues for learning approaches is the issue of tractable learning with large state-action spaces [2]. Many early approaches to policy learning for dialogue systems used small state spaces and action sets, and concentrated on only limited policy learning experiments (for example, type of confirmation, or type of initiative [12].) In recent years, however, resarchers have started to use a variety of dimensionality reduction methods, amongst them Linear Function Approximation [7], Hierarchical RL [13], and Summary POMDPs [14]. Use of these methods has allowed researchers to learn policies for complete dialogue action sets (i.e. all possible dialogue acts available to the system), often using complex state representations (e.g. dialogue history in addition to filled/confirmed slots) rather than limited action choices [7]. State generalization techniques also allow previously unseen dialogue situations to be dealt with robustly.

2.2. Corpora and user simulations

One critical concern for such data-hungry learning approaches is the development of appropriate dialogue corpora for training and testing. The COMMUNICATOR dataset [15] is the largest available corpus of human-machine dialogues, and has been further annotated with dialogue contexts [16]. This corpus has been extensively used for training and testing dialogue managers [7] and user simulations [17]. However useful it has been, COMMUNICATOR is restricted to information seeking dialogues (in the flight booking domain) for a small number of user-provided constraints (e.g. destination city, departure date), and the original annotations are somewhat lacking in consistency. Thus, new data collections, focussing on different aspects of dialogue and different genres (e.g. tutorial dialogue) are a top priority for the research community as a whole [18]. It is vital that newly collected corpora include reward annotations to allow reinforcement techniques.

Even when large data sets are available, in exploratory learning of dialogue strategies it is rarely the case that enough training data is available to sufficiently explore the vast space of possible dialogue states and strategies, and the best strategy may often not even be present in a given dataset. A promising approach is to use existing (small) corpora to train stochastic models for simulating user behavior, i.e. the way users interact with the system in order to accomplish their goals. Using real users would require much more time and effort – for example in current RL work tens of thousands of dialogues are often required for training. For these reasons dialogue simulation is often re-

quired to expand the existing dataset and human-machine spoken dialogue stochastic modelling and simulation has become a lively research field in its own right [3, 4, 5, 19, 20, 21, 22].

The validity of the assumption that performance in a simulated dialogue is an accurate indication of how the system would perform with real users is still an open research question (though see [23]) and reliable objective metrics for the evaluation of user simulations are a matter of debate [22, 24, 25, 26].

3. Main research areas - State of the Art

3.1. Policy learning with MDPs

In the MDP formalism, a discrete-time stochastic system interacting with its environment through actions is described by a finite or infinite number of states $\{s_i\}$ in which a given number of actions $\{a_j\}$ can be performed. Each state-action pair is associated with a transition probability $\mathcal{T}^a_{ss'}$: the probability of stepping from state s at time t to state s' at time $t+1$ after having performed action a when in state s. This transition is also associated with a reinforcement signal (or reward) r_{t+1} describing how good the result of action a was when performed in state s. If we denote the expected immediate reward by $\mathcal{R}^a_{ss'}$, the couple $\{\mathcal{T}, \mathcal{R}\}$ defines the dynamics of the system.

To control a system described as an MDP, one then needs a strategy or policy π mapping all states to actions: $\pi(s) = P(a|s)$ (or $\pi(s) = a$ if the strategy is deterministic). In this framework, a RL agent is a system aiming at optimally mapping states to actions, i.e. finding the best strategy π^* so as to maximize an overall reward R which is a function (most often a weighted sum) of all the immediate rewards. If the transition probabilities are known, an analytical solution can be computed by dynamic programming, otherwise the system has to learn the optimal strategy by a trial-and-error process [2]. In the most challenging cases, actions may affect not only the immediate reward, but also the next situation and, through that, all subsequent rewards. Trial-and-error search and delayed rewards are the two main features of RL.

In this framework, task-oriented human-machine dialogue can be modelled as a turn-taking process in which a human user and a Dialogue Manager (DM) exchange information through different channels, processing speech inputs and outputs (ASR, TTS ...). In the specific case of dialogue management policy learning, it is the DM strategy that has to be optimized and the DM will therefore be the learning agent. The learning environment includes everything but the DM: the human user, the communication channels (ASR, TTS ...), and any external information source (database, sensors etc.). In this context, at each dialogue turn t the DM has to choose an action a_t according to its current policy π_t and its internal state s_t, so as to attempt to complete the task it has been designed for. The dialogue state[1] is often built in a manner which reflects the amount of information received from the environment during the t previous turns (e.g. the filled and confirmed information "slots" such as departure date). Several dialogue action types have to be introduced, such as greetings, constraining questions, confirmations, data presentation etc. Performing these results in a response from the DM's environment (real user speech input, simulated user dialogue acts, returned database records etc.), considered as an observation o_t, which leads to a DM internal state update (s_{t+1}). In the MDP paradigm, it is assumed that a direct mapping between observations and states can be found (see section 3.2 for a discussion of partial observability).

In the MDP formalism, a reinforcement signal r_{t+1} is also required. In [1], the reinforcement signal was intuitively set to a weighted sum of objective and subjective parameters linked to the task. In [3] it was proposed to use actions' contributions to the user's satisfaction. Although this seems very subjective, the PARADISE study has shown that such a reward could be approximated by a linear combination of objective measures such as the duration of the dialogue, the ASR performances, and task completion [28]. Within this framework, it has been shown that effective dialogue policies can be learned [1, 3, 4, 7, 5, 6, 23] rather than crafted "by-hand".

3.2. Policy learning with POMDPs

MDPs provide a principled framework for Reinforcement Learning, but do not allow for uncertainty about observations of the environment. In dialogue systems, uncertainty is undeniably present, even if speech recognition is perfect. As well as the multiple possible speech recognition hypotheses that should be taken into account, there are semantic and pragmatic ambiguities also leading to uncertainties about the user's goals and intentions [29, 14] which mean that a partially observable view of dialogue context is much more accurate than the standard MDP model. In POMDPs, the dialogue policy is based not on the context at time t, but on the distribution over possible contexts at time t. The optimal system dialogue act to perform at time t then automatically takes account of the uncertainty in the context. The use of POMDPs in dialogue systems has so far been limited [14, 30, 31, 10], since the inference algorithms needed to choose a system action are computationally complex [29]. However, this is an extremely attractive model, offering graceful error handling and recovery, if the the tractability issues can be overcome.

3.3. User simulation techniques

As discussed previously, the small amount of data available for learning and testing dialogue strategies has led to a new field of research: human-machine stochastic dialogue modelling and simulation [19, 3, 4, 20, 21]. Among state-of-the-art simulation methods, one can distinguish between state-transition or global methods, like those proposed in [3], and methods based on modular simulation environments as described in [19, 4, 20, 21]. The first type of method is very task-dependent, as is the mixed method proposed in [4]. The second type of method integrates models of each component of a SDS including the speech processing systems but also the user. User modelling for spoken interaction is currently an important domain of investigation within the broad field of research on SDS and statistical methods are of particular interest.

Most of the simulation methods suppose that the interaction can be modelled at the intention level. From this, several statistical methods manipulating intentions to generate user behaviors have been proposed in the literature. First, the N-Gram model proposed in [32] assumes that the next user's utterance u_t can be inferred from the history h_t of the interaction using a set of conditional probabilities $P(u_t|h_t)$. Since a model conditioning the probability on all possible histories was impossible to train given the available amount of data, a bigram was used and only the previous system sentence is taken into account ($P(u_t|sys_{t-1})$). In [19], the same authors propose to use a set of general conditional probabilities to generate more

[1] Note that a common misunderstanding is that the Markov Property constrains models of dialogue state to exclude the dialogue history. However, it is possible to use variables in the current state which represent features of the dialogue history [27, 7, 6].

task-independent user behaviors. However, the one-step memory of the model makes it somewhat inconsistent. A similar method is used in [33, 34], but the authors propose to add a memory to the user so as to ensure that generated dialogues are coherent according to the user's goal and knowledge about the task. The N-gram idea has also been investigated further in [22] where longer histories are studied as well as richer dialogue state representations. In [4], a graph-based method is proposed combining rule-based decisions to model goal-directed user behavior while a stochastic model is used to generate the conversational behavior. Bayesian networks have also been proposed in [5, 35]. Recent work has also led to approaches based on Hidden Markov Models [36].

A large number of different approaches have also contributed to the development of user model evaluation methods [37, 24, 26]. This is very much an open domain of research.

3.4. Context-sensitive speech recognition

A related area of research is that on exploiting high-level contextual features in speech recognition, e.g. [38]. For example, a 60% reduction in the ASR error rate has been achieved for a complex task-based dialogue system using memory-based learning and dialogue context [39]. This research shows that a combination of low-level and context-based features is critical in improving dialogue system performance.

In general, contextual feedback can be accomplished by allowing higher level dialogue processing modules to select from lists of hypotheses provided by the lower level modules. This is sometimes called a reranking approach [40]. The effect is that the highest ranking hypotheses after reranking are those which are consistent with both the constraints from the lower levels and plausibility constraints from higher level modules.

3.5. Trainable Natural Language Generation

"Trainable" Natural Language Generation (NLG) [41, 42] is a recent approach where automatic techniques are used to train NLG modules, or to adapt them to specific domains and/or types of user. In SPaRKy [42], for example, candidate sentence plans are generated and then ranked. This process outputs a set of text-plan trees, which consist of speech acts to be communicated, and the rhetorical relations between them, which are then sent to a surface realizer. The candidate sentence plans are generated by an ordered set of clause-combining operations, while the stochastic part of the process is limited to training the sentence plan ranker, which uses rules learned from a labelled set of examples, using the RankBoost algorithm.

4. Conclusions

Theoretical and technological advances in several fields of human-machine spoken communication have made possible the recent investigation of spoken dialogue systems as statistical systems, using computational learning techniques. Although the necessary data to train a wide range of systems is not yet available, we have described several successful approaches to dialogue policy learning, and simulation techniques that will allow the expansion of existing small datasets to many unseen situations. Current results in the field of dialogue systems optimization using data-driven methods are very promising. However, this is still a relatively new and open research area, providing many opportunities for new theoretical advances and applications. The area is also generating new research topics, such as evaluation methods for user simulations. Further related areas are statistical parsing for speech and statistical semantic interpretation, see e.g. [43], and their relation to dialogue context models.

5. Acknowledgements

Oliver Lemon thanks the EPSRC (grant no. EP/E019501/1) and the TALK project (EC IST 507802). Olivier Pietquin thanks the SIMILAR network of excellence, the Belgian Walloon Region and the French Lorraine Region. The authors thank James Henderson, Kallirroi Georgila, and Steve Young.

6. References

[1] E. Levin, R. Pieraccini, and W. Eckert, "Learning dialogue strategies within the markov decision process framework," in *Proc. ASRU'97*, December 1997.

[2] R.S. Sutton and A.G. Barto, *Reinforcement Learning : An Introduction*, MIT Press, ISBN : 0-262-19398-1, 1998.

[3] S. Singh, M. Kearns, D. Litman, and M. Walker, "Reinforcement learning for spoken dialogue systems," in *Proc. NIPS'99*, 1999.

[4] K. Scheffler and S. Young, "Corpus-based dialogue simulation for automatic strategy learning and evaluation," in *Proc. NAACL Workshop on Adaptation in Dialogue Systems*, 2001.

[5] O. Pietquin and T. Dutoit, "A probabilistic framework for dialog simulation and optimal strategy learning," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 2, pp. 589–599, March 2006.

[6] M. Frampton and O. Lemon, "Learning more effective dialogue strategies using limited dialogue move features," in *Proceedings of ACL*, 2006.

[7] 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.

[8] J. Williams, P. Poupart, and S. Young, "Partially observable markov decision processes with continuous observations for dialogue management," in *Proceedings of the SigDial Workshop (SigDial'06)*, 2005.

[9] S. Young, "Using POMDPs for dialog management," in *Proceedings of the 1st IEEE/ACL Workshop on Spoken Language Technologies (SLT'06)*, 2006.

[10] JD Williams and SJ Young, "Partially Observable Markov Decision Processes for Spoken Dialog Systems," *Computer Speech and Language*, vol. 21, no. 2, pp. 231–422, 2007.

[11] Tim Paek and David Maxwell Chickering, "The Markov Assumption in spoken dialogue management," in *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*, 2005.

[12] Satinder Singh, Diane Litman, Michael Kearns, and Marilyn Walker, "Optimizing dialogue management with reinforcement learning: Experiments with the NJFun system," *Journal of Artificial Intelligence Research (JAIR)*, 2002.

[13] Oliver Lemon, Xingkun Liu, Daniel Shapiro, and Carl Tollander, "Hierarchical Reinforcement Learning of Dialogue Policies in a development environment for dialogue

systems: REALL-DUDE," in *Proceedings of Brandial, the 10th SemDial Workshop on the Semantics and Pragmatics of Dialogue, (demonstration systems)*, 2006.

[14] Jason Williams and Steve Young, "Scaling Up POMDPs for Dialog Management: The "Summary POMDP" Method," in *ASRU 05, Automatic Speech Recognition and Understanding Workshop*, 2005.

[15] Marilyn A. Walker, Rebecca J. Passonneau, and Julie E. Boland, "Quantitative and qualitative evaluation of DARPA Communicator spoken dialogue systems," in *Proc. ACL*, 2001, pp. 515–522.

[16] 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.

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

[18] G. Andreani, G. Di Fabbrizio, M. Gilbert, D. Gillick, D. Hakkani-Tür, and O. Lemon, "Let's DiSCoH: collecting an annotated open corpus with dialogue acts and reward signals for natural language helpdesks," in *IEEE/ACL Spoken Language Technology*, 2006.

[19] E. Levin, R. Pieraccini, and W. Eckert, "A stochastic model of human-machine interaction for learning dialog strategies," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 1, pp. 11–23, 2000.

[20] R. López-Cózar, A. de la Torre, J. Segura, and A. Rubio, "Assesment of dialogue systems by means of a new simulation technique," *Speech Communication*, vol. 40, no. 3, pp. 387–407, May 2003.

[21] O. Pietquin, "A probabilistic description of man-machine spoken communication," in *Proc. ICME'05*, July 2005.

[22] Kallirroi Georgila, James Henderson, and Oliver Lemon, "Learning User Simulations for Information State Update Dialogue Systems," in *Eurospeech*, 2005.

[23] Oliver Lemon, Kallirroi Georgila, and James Henderson, "Evaluating Effectiveness and Portability of Reinforcement Learned Dialogue Strategies with real users: the TALK TownInfo Evaluation," in *IEEE/ACL Spoken Language Technology*, 2006.

[24] Verena Rieser and Oliver Lemon, "Cluster-based user simulations for learning dialogue strategies and the super evaluation metric," in *Proceedings of Interspeech/ICSLP*, 2006.

[25] Jost Schatzmann, Karl Weilhammer, Matt Stuttle, and Steve Young, "A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies," *The Knowledge Engineering Review*, vol. 21, pp. 97–126, 2006.

[26] K. Georgila, J. Henderson, and O. Lemon, "User simulation for spoken dialogue systems: Learning and evaluation," in *Proc. Interspeech'06*, September 2006.

[27] Oliver Lemon, Kallirroi Georgila, James Henderson, Malte Gabsdil, Ivan Meza-Ruiz, and Steve Young, "D4.1: Integration of Learning and Adaptivity with the ISU approach," Tech. Rep., TALK Project, 2005.

[28] M. Walker, D. Litman, C. Kamm, and A. Abella, "Paradise: A framework for evaluating spoken dialogue agents," in *Proc. of the 35th Annual Meeting of the Association for Computational Linguistics*, 1997, pp. 271–280.

[29] B. Zhang, Q. Cai, J. Mao, and B. Guo, "Planning and acting under uncertainty: A new model for spoken dialogue system," in *Proc 17th Conf on Uncertainty in AI*, Seattle, 2001.

[30] Jason Williams, Pascal Poupart, and Steve Young, " Partially Observable Markov Decision Processes with Continuous Observations for Dialogue Management," in *Proceedings of the 6th SigDial Workshop on Discourse and Dialogue*, 2005.

[31] Jason Williams, Pascal Poupart, and Steve Young, "Factored Partially Observable Markov Decision Processes for Dialogue Management," in *4th Workshop on Knowledge and Reasoning in Practical Dialog Systems, International Joint Conference on Artificial Intelligence (IJCAI)*, 2005.

[32] W. Eckert, E. Levin, and R. Pieraccini, "User modeling for spoken dialogue system evaluation," in *Proc. ASRU'97*, December 1997.

[33] Olivier Pietquin and Steve Renals, "Asr system modeling for automatic evaluation and optimization of dialogue systems," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2002)*, Orlando, (USA, FL), May 2002.

[34] O. Pietquin, "Consistent goal-directed user model for realisitc man-machine task-oriented spoken dialogue simulation," in *Proc. ICME'06*, July 2006.

[35] O. Pietquin and T. Dutoit, "Dynamic bayesian networks for nlu simulation with applications to dialog optimal strategy learning," in *Proc. ICASSP'06*, May 2006.

[36] Heriberto Cuayáhuitl, Steve Renals, Oliver Lemon, and Hiroshi Shimodaira, "Human-computer Dialogue Simulation using Hidden Markov Models," in *ASRU*, 2005.

[37] J. Schatzmann, K. Georgila, and S. Young, "Quantitative evaluation of user simulation techniques for spoken dialogue systems," in *Proc. SIGdial'05*, September 2005.

[38] R. Jonson, "Dialogue Context-Based Re-ranking of ASR Hypotheses," in *Proceedings IEEE 2006 Workshop on Spoken Language Technology*, 2006.

[39] Malte Gabsdil and Oliver Lemon, "Combining acoustic and pragmatic features to predict recognition performance in spoken dialogue systems," in *Proceedings of ACL-04*, 2004, pp. 344–351.

[40] Ananlada Chotimongkol and Alexander I. Rudnicky, "N-best Speech Hypotheses Reordering Using Linear Regression," in *Proceedings of EuroSpeech 2001*, 2001, pp. 1829–1832.

[41] M. Walker, O. Rambow, and M. Rogati, "Spot: A trainable sentence planner," in *In Proc. of the NAACL,*, 2001.

[42] Amanda Stent, Rashmi Prasad, and Marilyn Walker., "Trainable sentence planning for complex information presentation in spoken dialog systems," in *Association for Computational Linguistics*, 2004.

[43] Y He and SJ Young, "Semantic Processing using the Hidden Vector State Model," *Computer Speech and Language*, vol. 19, no. 1, pp. 85–106, 2005.