TOWARDS NATURAL SPOKEN INTERACTION WITH ARTIFICIAL INTELLIGENT SYSTEMS

Stefan Ultes

Department of Engineering, University of Cambridge
su259@cam.ac.uk

Abstract: The skill of natural spoken interaction is crucial for artificial intelligent systems. To equip these systems with this skill, model-based statistical dialogue systems are essential. However, this goal is still far from reach. In this paper, the basics of statistical spoken dialogue systems, which play a key role in natural interaction, are presented. Furthermore, I will outline two principal aspects and argue why those are important to achieve natural interactions.

1 Introduction

Artificial intelligent systems are becoming prevalent in our everyday lives. Equipping such system with the ability to have a natural spoken interaction with humans is very important as machines and humans will soon closely collaborate in many tasks where efficient communication is essential. Natural spoken language is one of the most efficient means of communication and will play a key role.

To realise this, state-of-the-art research focusses on statistical spoken dialogue systems (SDS) which utilize machine learning algorithms thus disconnecting the system from the abilities of a human designer. Especially reinforcement learning (RL), where the system learns to behave to optimise a given objective, allows the system to learn behaviour which is specific to human-machine interaction instead of imitating human-human behaviour[1]. This is relevant as the expected behaviour of an artificial system might be different from human behaviour.

2 Statistical Spoken Dialogue Systems

Statistical spoken dialogue systems [2, 3] are based on partially observable Markov decision processes (POMDP) [4]. The key idea is to learn the optimal behaviour represented by the policy π and a delayed reward r in an environment where the current state cannot be determined exactly. Instead, a probability distribution over all possible states s is maintained called the belief b which is used as the input to the policy. To update the belief in a statistical SDS, the uncertainty of the user input is modelled on all input levels having n-best hypotheses for the ASR result and their semantic interpretations as shown in Figure 1. This observation o is then used to update the belief state. Hence, the new belief b' for state s' is defined as

$$b'(s') = P(o'|s',a) \cdot \sum_{s} P(s'|s,a) \cdot b(s) , \tag{1}$$

where a is the last system action. This general belief update is already intractable to maintain for small problems. Hence, in practice, approximative approaches have been proposed like

[1]Conventional approaches that analyse the behaviour of the system and whether it is aligned with human expectation rely on user studies and data that are biased by human designers. Reinforcement Learning (RL), though, allows to directly learn adequate system behaviour which is aligned with human expectations by definition. This has already shown to produce new and unexpected behaviour [1].

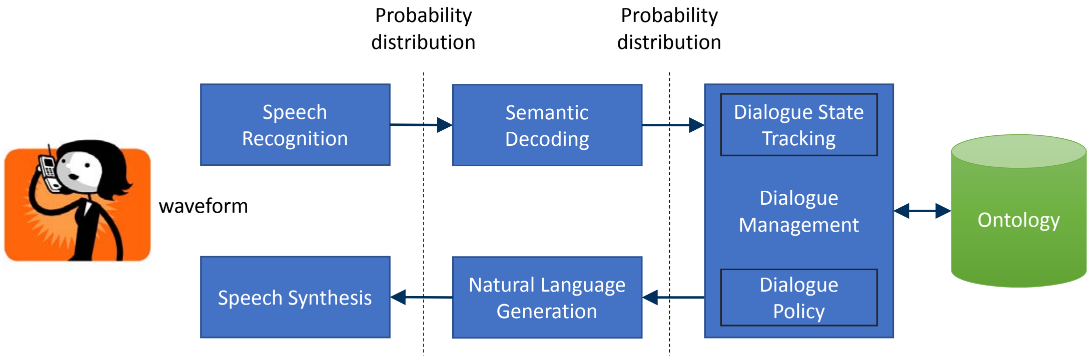

Figure 1 – Architecture of a statistical spoken dialogue system where probability distributions of word hypotheses or dialogue acts, respectively, are used to carry over the uncertainty of the user input.

the hidden information state approach [5] or Bayes nets [6]. More recently, discriminative approaches have been applied to track the belief state [7, 8, 9].

The system behaviour is then defined by the policy π which takes the current belief state as input to select the next system action a:

$$\pi(b) = a\ . \tag{2}$$

In order to find the optimal policy π^*, reinforcement learning is used. Unfortunately, the native RL approach within the POMDP formalism is intractable in practice. However, discrete-space POMDPs may be viewed as continuous Markov decision processes. By that, a variety of algorithms may be to applied to this simpler problem.

In general, reinforcement learning is used in a sequential decision-making process where a decision-model (the policy π) is trained based on sample data and a potentially delayed objective signal (the reward r) [10]. Implementing the Markov assumption, the policy selects the next action $a \in A$ based on the current system belief state b to optimise the accumulated future reward R_t at time t:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}\ . \tag{3}$$

Here, k denotes the number of future steps, γ a discount factor and r_τ the reward at time τ.

The Q-function models the expected accumulated future reward R_t when taking action a in belief state b and then following policy π:

$$Q^\pi(b,a) = E_\pi[R_t | b_t = b, a_t = a]\ . \tag{4}$$

One way of defining a policy is by utilizing the Q-function and select the action a as next system action with the maximum Q value:

$$\pi(b) = \operatorname*{argmax}_a Q^\pi(b,a)\ . \tag{5}$$

For most real-world problems, finding the exact optimal Q-values is not feasible. Instead, the Q-function is usually approximated. Within statistical SDS, various methods have been successfully applied for this [11, 12] including Gaussian processes [13] and Deep Q-Networks [14, 15].

More recently, policy-based methods have also been successfully applied to spoken dialogue systems where the parameters of a policy are directly learned [16, 17].

In order to train any of the above models, two key issues have to be resolved: the algorithm must learn to act in a large belief state and action space and training data is needed. To alleviate the large space problem, the optimisation in performed in a reduced space—the summary space—which is built according to heuristics. In this summary space, a summary action

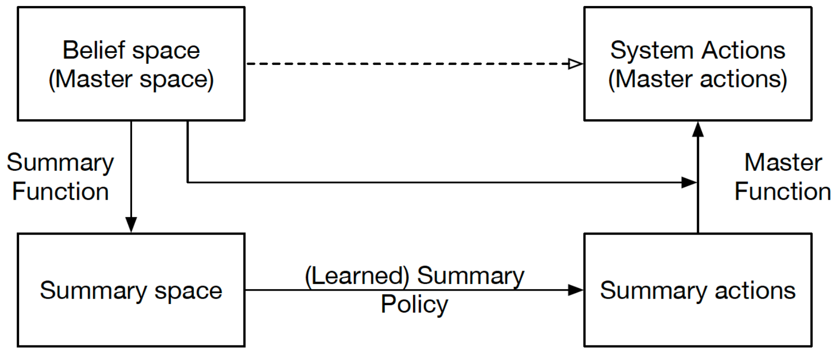

Figure 2 – Master and summary space used for learning dialogue behaviour in statistical spoken dialogue systems.

is selected which is then mapped back to an actual master action in master space as shown in Figure 2.

To get training dialogues, two methods have been proven to be useful. For development and bootstrapping, a user simulator may be used [18]. It is modelled to act as a user would do to create reasonable dialogues. The other option is to train with real people [19]. This is only possible, though, if the number of training dialogues needed is relatively small and intelligent means for collecting the reward are applied [20, 21].

3 Natural Interaction

To to equip an artificial system with the skill of natural spoken communication, the corresponding SDS must be able to understand the user input by putting it into the context of the whole interaction as well as to produce adequate responses. Hence, a system must be able to process complex dialogue structures and behave in a way which is perceived as natural by the user.

Even though there have been many contributions to the state-of-the-art in goal-oriented statistical spoken dialogue systems recently, the complexity of possible dialogue structures has remained rather limited. Instead, recent work proposed new RL algorithms [17, 15], new state models [22, 23], or new system models [24, 25, 26].

Recent statistical SDSs that use RL [2, 27] are model-based where the dialogue model controls the complexity of dialogue structures that can be processed by the system[2]. As it is common for RL-base systems, the system behaviour is defined by the objective function. Thus, to increase the naturalness of the dialogues, new dialogue models are needed allowing for more complex dialogue structures and dialogue objectives need to be investigated which allow for learning more natural system behaviour beyond task success.

The Dialogue Model The role of the dialogue model is to define the structure and the internal links of the dialogue state as well as the set of available system and user acts (i.e., the semantic representation of the user input utterance). It further defines the abstraction layer interfacing the background knowledge base. Most current models are build around *domains* which encapsulate all relevant information as a section of the dialogue state that belongs to a given topic, e.g., finding a *restaurant* or *hotel*. However, the resulting flat domain-centred state that is widely used is not intuitive to model more complex dialogue structures like relations (e.g. 'I am looking for a hotel and a restaurant in the same area', 'I need a taxi to the station in time to catch

[2]Model-free approaches like end-to-end generative networks [26, 28] have interesting properties (e.g., they only need text data for training) but they still seem to be limited in terms of dialogue structure complexity (not linguistic complexity) in cases where content from a structured knowledge base needs to be incorporated. Approaches where incorporating this information is learned along with the system responses based on dialogue data [29] seem hard to scale. Furthermore, it seems to be counter-intuitive to learn this solely from dialogue data.

the train'), multiple entities (e.g., two restaurants) or connecting a set of objects (e.g., adding several participants to a calendar entry). More realistic dialogue models are needed along with RL algorithms that are able to deal with this added complexity.

The Dialogue Objective The dialogue objective is used in RL-based dialogue systems to guide the learning process distinguishing good from bad system behaviour. The current standard objective for goal-oriented dialogue systems is task success. While it is undoubtedly most important, is fails at capturing aspects of natural interaction: there are many policies that lead to successful dialogues, but what is the subset of policies that lead to an interesting, satisfying, funny, natural, polite, etc. interaction? And what implications can be drawn for the system responses in order to achieve this? Novel methods are needed both to incorporate these objectives in a feasible way and to investigate the implications on the system response.

Even though the model and the objective are both important for natural interaction, the dialogue model plays a core role. It does not only define the possible dialogue structures but also all possibilities how the system can express itself. And for finding feasible dialogue objectives in the increased complexity induced by the dialogue model, a good starting point are methods that build on active learning [21], automatic estimation of the objective [30] or reward balancing [31].

4 Conclusion

Statistical spoken dialogue systems are essential for achieving the goal of equipping an artificial intelligent agent with the skill of natural spoken interaction. While the complexity of dialogue structures and the dialogue objective play key roles in natural interaction, there are more aspects that are required to be realised like the linguistic variability of the system response [32, 25], integrating complex knowledge bases [33], or adapting the system behaviour to the user state [34, 35, 36].

References

[1] SILVER, D., A. HUANG, C. J. MADDISON, A. GUEZ, L. SIFRE, G. VAN DEN DRIESSCHE, J. SCHRITTWIESER, I. ANTONOGLOU, V. PANNEERSHELVAM, M. LANCTOT ET AL.: *Mastering the game of go with deep neural networks and tree search*. *Nature*, 529(7587), pp. 484–489, 2016.

[2] YOUNG, S. J., M. GAŠIĆ, B. THOMSON, and J. D. WILLIAMS: *POMDP-based statistical spoken dialog systems: A review*. *Proceedings of the IEEE*, 101(5), pp. 1160–1179, 2013.

[3] ULTES, S., L. M. ROJAS-BARAHONA, P.-H. SU, D. VANDYKE, D. KIM, I. CASANUEVA, P. BUDZIANOWSKI, N. MRKŠIĆ, T.-H. WEN, M. GAŠIĆ, and S. J. YOUNG: *Pydial: A multi-domain statistical dialogue system toolkit*. In *ACL Demo*. Association of Computational Linguistics, 2017.

[4] KAELBLING, L. P., M. L. LITTMAN, and A. R. CASSANDRA: *Planning and acting in partially observable stochastic domains*. *Artificial Intelligence*, 101(1-2), pp. 99–134, 1998.

[5] YOUNG, S. J., M. GAŠIĆ, S. KEIZER, F. MAIRESSE, J. SCHATZMANN, B. THOMSON, and K. YU: *The hidden information state model: A practical framework for POMDP-*

based spoken dialogue management. *Computer Speech & Language*, 24(2), pp. 150–174, 2010.

[6] THOMSON, B. and S. J. YOUNG: *Bayesian update of dialogue state: A POMDP framework for spoken dialogue systems*. *Computer Speech & Language*, 24(4), pp. 562–588, 2010.

[7] HENDERSON, M., B. THOMSON, and S. J. YOUNG: *Deep neural network approach for the dialog state tracking challenge*. In *Proceedings of the SIGDIAL 2013 Conference*, pp. 467–471. 2013.

[8] HENDERSON, M., B. THOMSON, and S. J. YOUNG: *Robust dialog state tracking using delexicalised recurrent neural networks and unsupervised adaptation*. In *Spoken Language Technology Workshop (SLT), 2014 IEEE*, pp. 360–365. IEEE, 2014.

[9] MRKŠIĆ, N., D. Ó SÉAGHDHA, T.-H. WEN, B. THOMSON, and S. YOUNG: *Neural belief tracker: Data-driven dialogue state tracking*. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1777–1788. Association for Computational Linguistics, 2017. doi:10.18653/v1/P17-1163. URL `http://www.aclweb.org/anthology/P17-1163`.

[10] SUTTON, R. S. and A. G. BARTO: *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, USA, 1st edn., 1998. URL `http://portal.acm.org/citation.cfm?id=551283`.

[11] LEFEVRE, F., M. GAŠIĆ, F. JURČÍČEK, S. KEIZER, F. MAIRESSE, B. THOMSON, K. YU, and S. J. YOUNG: *k-nearest neighbor Monte-Carlo control algorithm for POMDP-based dialogue systems*. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 272–275. Association for Computational Linguistics, 2009.

[12] PIETQUIN, O., M. GEIST, and S. CHANDRAMOHAN: *Sample efficient on-line learning of optimal dialogue policies with kalman temporal differences*. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, p. 1878. 2011.

[13] GAŠIĆ, M. and S. J. YOUNG: *Gaussian processes for POMDP-based dialogue manager optimization*. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(1), pp. 28–40, 2014.

[14] PAPANGELIS, A. and Y. STYLIANOU: *Single-model multi-domain dialogue management with deep learning*. In *International Workshop for Spoken Dialogue Systems*. 2017.

[15] CASANUEVA, I., P. BUDZIANOWSKI, P.-H. SU, N. MRKŠIĆ, T.-H. WEN, S. ULTES, L. ROJAS-BARAHONA, S. YOUNG, and M. GAŠIĆ: *A benchmarking environment for reinforcement learning based task oriented dialogue management*. *arXiv preprint arXiv:1711.11023*, 2017.

[16] SU, P.-H., M. GASIC, N. MRKSIC, L. ROJAS-BARAHONA, S. ULTES, D. VANDYKE, T.-H. WEN, and S. YOUNG: *Continuously learning neural dialogue management*. *arXiv preprint arXiv:1606.02689*, 2016.

[17] SU, P.-H., P. BUDZIANOWSKI, S. ULTES, M. GASIC, and S. YOUNG: *Sample-efficient actor-critic reinforcement learning with supervised data for dialogue management*. In *SIGdial*, pp. 147–157. ACL, Saarbrücken, Germany, 2017.

[18] SCHATZMANN, J. and S. J. YOUNG: *The hidden agenda user simulation model. Audio, Speech, and Language Processing, IEEE Transactions on*, 17(4), pp. 733–747, 2009.

[19] GAŠIĆ, M., C. BRESLIN, M. HENDERSON, D. KIM, M. SZUMMER, B. THOMSON, P. TSIAKOULIS, and S. J. YOUNG: *On-line policy optimisation of Bayesian spoken dialogue systems via human interaction.* In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8367–8371. IEEE, 2013.

[20] SU, P.-H., D. VANDYKE, M. GAŠIĆ, D. KIM, N. MRKŠIĆ, T.-H. WEN, and S. J. YOUNG: *Learning from real users: Rating dialogue success with neural networks for reinforcement learning in spoken dialogue systems.* In *Interspeech*, pp. 2007–2011. ISCA, 2015.

[21] SU, P.-H., M. GAŠIĆ, N. MRKŠIĆ, L. ROJAS-BARAHONA, S. ULTES, D. VANDYKE, T. H. WEN, and S. YOUNG: *On-line active reward learning for policy optimisation in spoken dialogue systems.* In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 2431–2441. Association for Computational Linguistics, 2016.

[22] SCHULZ, H., J. ZUMER, L. E. ASRI, and S. SHARMA: *A frame tracking model for memory-enhanced dialogue systems. arXiv preprint arXiv:1706.01690*, 2017.

[23] LEE, S. and A. STENT: *Task lineages: Dialog state tracking for flexible interaction.* In *SIGDial*, pp. 11–21. ACL, Los Angeles, 2016.

[24] WEN, T.-H., D. VANDYKE, N. MRKŠIĆ, M. GAŠIĆ, L. ROJAS-BARAHONA, P.-H. SU, S. ULTES, and S. YOUNG: *A network-based end-to-end trainable task-oriented dialogue system.* In *EACL*, pp. 438–449. ACL, 2017.

[25] WEN, T.-H., Y. MIAO, P. BLUNSOM, and S. YOUNG: *Latent intention dialogue models.* In *ICML*, ICML'17. JMLR.org, 2017.

[26] SERBAN, I. V., A. SORDONI, Y. BENGIO, A. C. COURVILLE, and J. PINEAU: *Building end-to-end dialogue systems using generative hierarchical neural network models.* In *AAAI*, pp. 3776–3784. 2016.

[27] LEMON, O. and O. PIETQUIN: *Data-Driven Methods for Adaptive Spoken Dialogue Systems.* Springer New York, 2012. doi:10.1007/978-1-4614-4803-7.

[28] LI, J., W. MONROE, A. RITTER, M. GALLEY, J. GAO, and D. JURAFSKY: *Deep reinforcement learning for dialogue generation.* In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1192–1202. Association for Computational Linguistics, 2016.

[29] ERIC, M. and C. D. MANNING: *Key-value retrieval networks for task-oriented dialogue.* In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 37–49. Association for Computational Linguistics, Saarbrücken, Germany, 2017. URL `http://www.aclweb.org/anthology/W17-366`.

[30] ULTES, S., P. BUDZIANOWSKI, I. CASANUEVA, N. MRKŠIĆ, L. ROJAS-BARAHONA, P.-H. SU, T.-H. WEN, M. GAŠIĆ, and S. YOUNG: *Domain-independent user satisfaction reward estimation for dialogue policy learning.* In *Interspeech*, pp. 1721–1725. ISCA, 2017.

[31] ULTES, S., P. BUDZIANOWSKI, I. CASANUEVA, N. MRKŠIĆ, L. M. R. BARAHONA, P.-H. SU, T.-H. WEN, M. GASIC, and S. YOUNG: *Reward-balancing for statistical spoken dialogue systems using multi-objective reinforcement learning*. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 65–70. 2017.

[32] WEN, T.-H., M. GAŠIĆ, N. MRKŠIĆ, P.-H. SU, D. VANDYKE, and S. YOUNG: *Semantically conditioned lstm-based natural language generation for spoken dialogue systems*. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1711–1721. Association for Computational Linguistics, Lisbon, Portugal, 2015. URL `https://aclweb.org/anthology/D/D15/D15-1199`.

[33] PAPANGELIS, A., P. PAPADAKOS, M. KOTTI, Y. STYLIANOU, Y. TZITZIKAS, and D. PLEXOUSAKIS: *Ld-sds: Towards an expressive spoken dialogue system based on linked-data*. *arXiv preprint arXiv:1710.02973*, 2017.

[34] ULTES, S., T. HEINROTH, A. SCHMITT, and W. MINKER: *A theoretical framework for a user-centered spoken dialog manager*. In R. LÓPEZ-CÓZAR and T. KOBAYASHI (eds.), *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop*, pp. 241–246. Springer New York, New York, NY, 2011. doi:10.1007/978-1-4614-1335-6_24.

[35] ULTES, S.: *User-centred Adaptive Spoken Dialogue Modelling*. Dissertation, Ulm University, Ulm, 2015.

[36] ULTES, S., H. DIKME, and W. MINKER: *Dialogue Management for User-Centered Adaptive Dialogue*. In A. I. RUDNICKY, A. RAUX, I. LANE, and T. MISU (eds.), *Situated Dialog in Speech-Based Human-Computer Interaction*, pp. 51–61. Springer International Publishing, Cham, 2016. doi:10.1007/978-3-319-21834-2_5.