One Arm to Rule Them All: Online Learning with Multi-armed Bandits for Low-resource Conversational Agents [*]

Vânia Mendonça[✉1,2][0000−0001−5729−7608], Luísa Coheur[1,2][0000−0002−2456−5028], and Alberto Sardinha[1,2][0000−0002−5782−3142]

[1] INESC-ID, Lisboa, Portugal
[2] Instituto Superior Técnico, Lisboa, Portugal
{vania.mendonca,luisa.coheur,jose.alberto.sardinha}@tecnico.ulisboa.pt

Abstract. In a low-resource scenario, the lack of annotated data can be an obstacle not only to train a robust system, but also to evaluate and compare different approaches before deploying the best one for a given setting. We propose to dynamically find the best approach for a given setting by taking advantage of feedback naturally present on the scenario in hand (when it exists). To this end, we present a novel application of online learning algorithms, where we frame the choice of the best approach as a multi-armed bandits problem. Our proof-of-concept is a retrieval-based conversational agent, in which the answer selection criteria available to the agent are the competing approaches (arms). In our experiment, an adversarial multi-armed bandits approach converges to the performance of the best criterion after just three interaction turns, which suggests the appropriateness of our approach in a low-resource conversational agent.

Keywords: Online learning · Multi-armed bandits · Conversational agents.

1 Introduction

State of the art on several Natural Language Processing tasks is currently dominated by deep learning approaches. In the particular case of conversational agents, such deep approaches have been applied to either generate an answer from scratch - *generation-based* - or to find the best match among a collection of candidate answers - *retrieval-based* -, with some works combining both approaches. Focusing on retrieval-based conversational agents, current approaches

[*] This work was supported by: Fundação para a Ciência e a Tecnologia (FCT) under reference UIDB/50021/2020 (INESC-ID multi-annual funding), as well as under the HOTSPOT project with reference PTDC/CCI-COM/7203/2020; Air Force Office of Scientific Research under award number FA9550-19-1-0020; P2020 program, supervised by Agência Nacional de Inovação (ANI), under the project CMU-PT Ref. 045909 (MAIA). Vânia Mendonça was funded by an FCT grant, ref. SFRH/BD/121443/2016.

often make use of large amounts of annotated data and/or heavy computations [8,29][3], which may not be viable in real world *low-resource* scenarios (i.e. scenarios that are scarce in datasets annotated with the appropriateness of each answer to a certain input). An alternative that could be more appropriate to a low-resource scenario would be an agent based on shallow criteria (e.g., similarity measures [3]) to select an answer.

Consider an agent equipped with an arbitrary number of answer selection criteria (either shallow or pre-trained). Assuming that we do not know in advance which criterion is going to be the best for a given setting (i.e., domain and/or language), how can the agent dynamically prioritize the best criterion (if such criterion exists) without a previous evaluation on an annotated dataset? One way to tackle this challenge is to take advantage of user feedback at each interaction to assess which criterion is doing the best job, using, for instance, online learning. We thus frame the problem of choosing a selection criterion at each interaction as a *multi-armed bandits* problem. Under this online learning framework, each selection criterion is an *arm*, and our goal is to converge towards the performance of the best criterion. Each selection criterion is evaluated in an online fashion, by taking advantage of human feedback available at each user interaction.

Existing similar proposals frame the choice of a selection criterion at each interaction as a problem of prediction with expert advice [19,18]. Unlike multi-armed bandits, this framework assumes that there is a single optimal outcome based on which the competing approaches (experts) are evaluated. However, in a conversational agent scenario, there is no single appropriate answer; moreover, the user is not expected to give feedback to all the experts, as only the agent's final answer will be presented to the user.

Thus, to the best of our knowledge, our work is the first to frame the problem of converging to the best answer selection criteria as a multi-armed bandits problem in a retrieval-based conversational agent scenario, keeping a low-resource setting in mind. Our experimental results show that an adversarial multi-armed bandits approach is able to converge towards the performance of the best individual expert after just three interaction turns, suggesting that this may be adequate for a low-resource setting.

2 From Prediction with Expert Advice to Multi-armed Bandits

A problem of prediction with expert advice can be seen as an iterative game between a *forecaster* and the *environment*, in which the forecaster consults different sources (*experts*) to provide the best forecast [7]. At each time-step t, the forecaster consults the predictions made by a set of K experts (each associated with a weight ω_k), in the decision space $\mathcal{D}$. Considering these predictions, the forecaster makes its own prediction, $\hat{p}^t \in \mathcal{D}$. At the same time, the environment reveals an outcome y^t in the decision space $\mathcal{Y}$ (which may not be exactly the same as $\mathcal{D}$).

[3] See Boussaha *et al* [5] for a review of recent retrieval-based systems.

Prediction with expert advice works under the assumption that the forecaster learns its own loss, ℓ^t, and the loss of each expert, ℓ_k^t , after the environment's outcome is revealed. In our conversational agent scenario, this assumption does not hold, since there is no single optimal outcome (i.e., a single appropriate answer), but instead there may be several appropriate outcomes (or none at all) among the candidate answers. Moreover, in a real world scenario, the user is not expected to give feedback to all the experts' answers, as only the agent's final answer will be presented to the user. Thus, we consider a related class of problems, *multi-armed bandits*, in which the environment's outcome is unknown, and only the forecaster learns its own loss [21,12]. In this class of problems, one starts by attempting to estimate the means of the loss distributions for each expert (here called *arm*) in the first iterations (the exploration phase), and when the forecaster has a high level of confidence in the estimated values, one may keep choosing the prediction with the smallest estimated loss (the exploitation phase).

A popular online algorithm for *adversarial* multi-armed bandits is Exponential-weighting for Exploration and Exploitation (EXP3) [2]. At each time step t, the forecaster's prediction is randomly selected according to the probability distribution given by the weights $\omega_1^{t-1}, \ldots, \omega_K^{t-1}$ of each arm k:

$$p_k^t = \frac{\omega_k^{t-1}}{\sum_{k'=1}^{K} \omega_{k'}^{t-1}} \tag{1}$$

Since only the arm selected by the forecaster knows its loss, only the weight of that arm is updated, as follows:

$$\omega_k^t = \omega_k^{t-1} e^{-\eta \hat{\ell}_k^t} \tag{2}$$

where η is the learning rate, and $\hat{\ell}_k^t = \frac{\ell_k^t}{p_k^t}$, (with ℓ_k^t being the loss obtained by the chosen arm k).

As for *stochastic* multi-armed bandit problems, i.e., problems where the loss is randomly sampled from an unknown underlying distribution, a popular algorithm used is Upper Confidence Bound (UCB) [1]. At each time step t, UCB estimates the average loss for each prediction, as well as a confidence interval, and selects the arm k with the lowest confidence bound (rather than the prediction with lowest estimated loss), as follows:

$$k^t = \operatorname*{argmin}_k \{\hat{Q}(k) - \sqrt{\frac{2log(t)}{N(k)}}\} \tag{3}$$

where $N(k)$ is the counter for how many times the arm k was selected by UCB, and $\hat{Q}(k)$ is the estimated cost associated with the arm k. $\hat{Q}(k)$ is updated whenever k corresponds to the arm selected by the forecaster, k^t, as follows:

$$\hat{Q}(k) = \hat{Q}(k) + \frac{1}{N(k)+1}(\ell_k^t - \hat{Q}(k)) \tag{4}$$

3 Related Work

Online learning, and particularly the multi-armed bandits framework, has been relatively under-explored in conversational agents and dialog systems, despite the interactive nature of this field. Several works have applied some form of online learning to conversational agents or dialog systems, most of them based on the Reinforcement Learning (RL) framework [4]. RL has been mostly applied to task-oriented dialog systems [13,24,9,25], but it has also been proposed in the context of non-task oriented systems: Yu *et al* [28] use RL to select a response strategy among a fixed set of available strategies; Serban et al [23] use RL to select a response from an ensemble of both retrieval and generation-based dialog systems.

More recently, online frameworks based on bandits have also been used in conversational agents and dialog systems. Genevay *et al* [10] applied multi-armed bandits for user adaptation in a task-oriented spoken dialog system, using the UCB algorithm to choose the best source user from which to transfer relevant information to a target new user. Upadhyay *et al* [26] applied contextual bandits [27] to select a skill to respond to a user query, in a virtual assistant scenario. Liu *et al* [15] used contextual bandits to select an answer from a pool of candidates at each turn, given the conversation context, in a retrieval-based conversational agent. The work with the closest goal to ours is that of Mendonça *et al* [19,18], who combine multiple answer selection criteria under the framework of prediction with expert advice. This framework assumes that there is a single optimal outcome based on which the competing approaches (experts) are evaluated. However, in a conversational agent scenario, there is no single appropriate answer, and the user is not expected to give feedback to all the experts, as only the agent's final answer will be presented to the user. Moreover, the authors did not show whether their approach indeed converged to the best performing criterion. Our work addresses these shortcomings by framing the problem of dynamically converging to the best answer selection criterion as a multi-armed bandits problem. This framework has the potential to be more suitable to the scenario in hand, since it does not require feedback for all the selection criteria, nor does it assume a single correct outcome.

4 Proof-of-concept: Retrieval-based Conversational Agent with Multi-armed Bandits

4.1 Finding the Best Answer Selection Criteria

In our scenario, an agent receives a user request and searches for an answer in a collection of interactions. We follow a *retrieve and refine* strategy[4] [22], i.e., after having retrieved a set of candidates, the agent takes advantage of a set of criteria to select a more appropriate answer. There may be several criteria available, and we may not know *a priori* which one is the best. We frame the choice of the

[4] However, we are not using generation and/or deep learning.

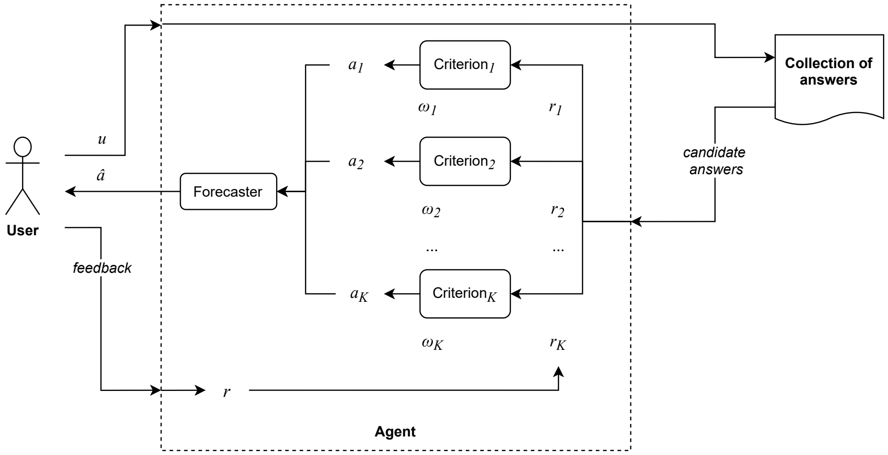

Fig. 1: Overview of the retrieval-based conversational agent scenario under the multi-armed bandits framework, for an interaction turn t.

best criterion as a multi-armed bandits problem, where each criterion is an arm associated with a weight ω_k. To learn the arms' weights, we apply the two online algorithms described in Section 2: EXP3 and UCB.

The learning process is shown in Fig. 1, and goes as follows: for each interaction turn t, the user sends a request u^t to the agent. The agent retrieves a set of candidate answers from its collection, and, from that set, each criterion (arm) $k = 1, \ldots, K$ chooses an answer a_k. The forecaster then chooses its answer $\hat{a}$ from the arms' answers $a_1, \ldots, a_K$, according to Eq. 1 for EXP3, and Eq. 3 for UCB. Then, the user evaluates $\hat{a}$ with a reward $r^t \in [0, 1]$. By setting each arm's loss as $\ell_k^t = -r_k^t$, the weight of the criterion k selected by the forecaster can be updated using the respective weight update rules (Eq. 2 for EXP3[5], and Eq. 4 for UCB[6]).

As our proof-of-concept scenario, we use Say Something Smart (SSS) [16], which has access to a collection of interactions (in the form of *trigger-answer* pairs) and selects an answer according to a combination of weighted criteria. In a first step (*retrieve*), given a user input, SSS selects a set of N candidate trigger-answer pairs[7] using Lucene [17]. Then, in a second step (*refine*), it applies the following criteria (which correspond to the arms in the multi-armed bandits setting):

- *Answer Frequency*: we consider the frequency of the candidate answers in the collection of interactions, following other systems based on the redundancy of the answer (such as the ones described in Lin [14] and Brill *et al* [6]);

[5] For EXP3, we rounded each arm's reward to an integer value, to avoid exploding weight values, and we set η to $\sqrt{8 \log \frac{K}{T}}$, following Mendonça *et al* [19].

[6] For UCB, we consider the estimated cost $\hat{Q}(k)$ as the "weight" for the arm k.

[7] We kept SSS's default configuration of $N = 20$ candidates.

```
P: Qual o custo do Cartão da Empresa e do Cartao de Pessoa Coletiva?
VG1: Qual é o custo do Cartão da Empresa e do Cartão Coletivo?
VG2: Qual é o custo do Cartão da Empresa e do Cartão Coletivo?
VUC: Quanto custa o cartão da empresa?
VUC: Quanto tenho de pagar pelo cartão de pessoa coletiva?
VIN: Qual o valor do Cartão da Empresa e do Cartão de Pessoa Coletiva?
VIN: Quanto custa o Cartão da Empresa e o Cartão de Pessoa Coletiva?
VIN: Qual o preço do Cartão da Empresa e do Cartão de Pessoa Coletiva?
R: Qualquer um dos cartões custa € 14,00 por unidade.

P: What is the cost of the Company Card and the Collective Person Card?
VG1: What is the cost of the Company Card and the Collective Card?
VG2: What is the cost of the Company Card and the Collective Card?
VUC: How much is the company card?
VUC: How much do I have to pay for the collective person card?
VIN: What is the value of the Company Card and the Collective Person Card?
VIN: How much is the Company Card and the Collective Card?
VIN: What is the price of the Company Card and the Collective Person Card?
R: Either of those cards costs € 14,00 per unit.
```

Fig. 2: Example entry from the AIA-BDE corpus, and its translation to English below [20].

- *Answer Similarity*: we consider that the answer can be a reformulation of the questions, following Lin [14]; thus, the similarity between the candidate answer and the user request is considered;
- *Trigger Similarity*: the similarity between the candidate trigger and the user request is considered.

Both Answer and Trigger similarity criteria use the Jaccard similarity measure. Note that, while in this experiment we use the criteria available in SSS, our approach is criterion-agnostic, thus it could be applied to any other set of criteria.

4.2 Obtaining User Feedback

We simulate user feedback using a reference corpus. At each learning step t, an interaction pair *trigger-answer* is selected from the reference corpus. The trigger tr is presented to the agent as being a user request. The agent retrieves a set of candidates from its collection of answers, and each arm k scores the different candidate answers, then choosing their highest scored answer as a_k. We simulate the user reward by measuring how well the answer $\hat{a}$ selected by the forecaster matches the reference answer, a^*, using the Jaccard similarity measure [11].

The corpus from which we built the agent's collection of interactions and the reference corpus was AIA-BDE[8] [20], a corpus of questions and answers in

[8] We use an updated version of the corpus reported by Oliveira *et al* [20], which includes more question variants for each answer.

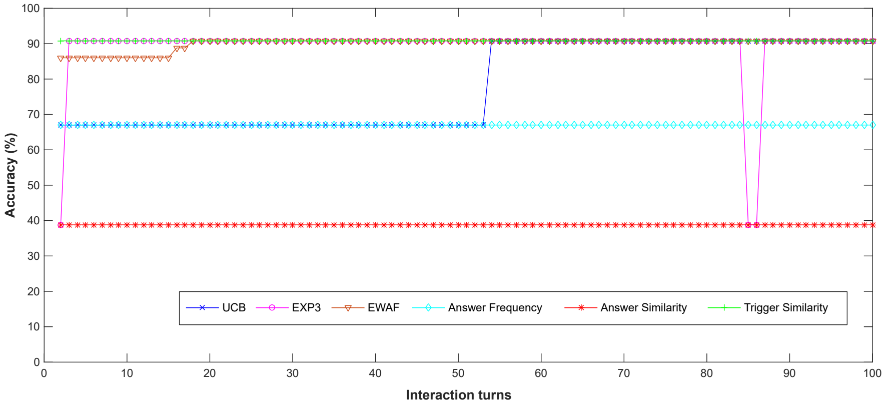

Fig. 3: Accuracy obtained by EWAF, EXP3, and UCB, as well as by each answer selection criterion.

Portuguese. In AIA-BDE, for each answer (R), there are several variants (VG1, VG2[9], VUC, VIN) of the corresponding question (P), as illustrated in Fig. 2.

In our experiment, we used all the pairs P-R (i.e., the gold question and answer pairs) as the agent's collection of interactions, and we used the question variations (VG1, VG2, VUC, VIN) paired with the answer R as the reference corpus. Out of these, we used 350 pairs to simulate a conversation with a user and learn the weights, and another 500 pairs to evaluate the performance of the weights learned at each iteration, in order to assess how well our online learning approach performs in the face of novel triggers (i.e., triggers that were not seen when learning the criteria's weights). We computed the accuracy, i.e., the percentage of iterations in which the agent chose the candidate answer that matched the input reference answer.

5 Experimental Results

In Fig. 3, we report the accuracy (%) of each multi-armed bandits algorithm, as well as each individual criterion, and we also compare our multi-armed bandits approach to that of Mendonça *et al* [19,18], who used Exponentially Weighted Average Forecaster (EWAF), a popular algorithm for prediction with expert advice [7]. For clarity, we only report up to 100 learning interaction turns, since the performance for each algorithm remains the same from then on.

Our first research question is whether any of the proposed multi-armed bandits approaches converges to the best criterion. As shown in Fig. 3, the perfor-

[9] VG1 and VG2 were obtained by translating P to English and back to Portuguese using the Google Translate API, once and twice, respectively [20]. Thus, duplicates, such as the one in Fig. 2, may occur.

mance of both EXP3 and UCB converges to that of the best answer selection criterion (which is, by far, Trigger similarity), similarly to the EWAF baseline.

Moreover, we investigate after how many interaction turns are needed for its performance to get close to the performance of the best selection criterion. EXP3 matches the performance of the best selection criteria from the very start (three interaction turns) until the end, with the exception of iterations 85-86 (in which EXP3 gave a greater weight to *Answer Frequency*).

On the other hand, UCB only gets closer to the best criterion at 54 interaction turns, thus taking longer to converge than the EWAF baseline, which converges after 18 interaction turns. A factor that may contribute to this difference in performance between EXP3 and UCB is that the latter assumes that the loss function is randomly sampled from a fixed unknown underlying distribution (which is not our case, since our reward function does not follow a fixed distribution), while EXP3 makes no such assumption. This outcome suggests that EXP3 may be a more adequate choice of algorithm for a conversational agent scenario, especially in a low-resource setting.

6 Conclusions and Future Work

In this work, we addressed a scenario where several approaches can be used and there is no gold data to properly evaluate them before deployment. We proposed an online learning approach based on the multi-armed bandits framework, and tested it on a retrieval-based conversational agent that relies on a number of criteria to select an answer. Our goal was to dynamically converge to the performance of the best answer selection criterion as the agent interacts with the user, taking advantage of their feedback, instead of evaluating each criteria *a priori*. In our experiment, in which we simulated the user feedback using a reference corpus composed of gold interaction pairs, the performance of the adversarial multi-armed bandits approach immediately matches that of the best performing selection criterion, which suggests this may be an adequate approach for a low-resource setting.

As for future work, we intend to expand this experiment by considering other answer selection criteria, as well as alternative loss functions.

References

1. Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multiarmed bandit problem. Machine Learning **47**(2-3), 235–256 (2002). https://doi.org/10.1023/A:1013689704352
2. Auer, P., Cesa-Bianchi, N., Freund, Y., Schapire, R.E.: Gambling in a rigged casino: the adversarial multi-armed bandit problem. In: Annual Symposium on Foundations of Computer Science - Proceedings. pp. 322–331 (1995). https://doi.org/10.1109/sfcs.1995.492488
3. Banchs, R.E., Li, H.: Iris: A chat-oriented dialogue system based on the vector space model. In: Proceedings of the ACL 2012 System Demonstrations. pp. 37–42. ACL '12, Association for Computational Linguistics, Stroudsburg, PA, USA (2012), http://dl.acm.org/citation.cfm?id=2390470.2390477

4. Biermann, A.W., Long, P.M.: The Composition of Messages in Speech-Graphics Interactive Systems. In: International Symposium on Spoken Dialogue. pp. 97–100 (1996), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.55.721{&}rep=rep1{&}type=pdf
5. Boussaha, B.E.A., Hernandez, N., Jacquin, C., Morin, E.: Deep Retrieval-Based Dialogue Systems: A Short Review. Tech. rep. (2019), http://arxiv.org/abs/1907.12878
6. Brill, E., Dumais, S., Banko, M.: An analysis of the askmsr question-answering system. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10. p. 257–264. EMNLP '02, Association for Computational Linguistics, USA (2002). https://doi.org/10.3115/1118693.1118726, https://doi.org/10.3115/1118693.1118726
7. Cesa-Bianchi, N., Lugosi, G.: Prediction, Learning and Games. Cambridge University Press (2006)
8. Chen, Q., Wang, W.: Sequential neural networks for noetic end-to-end response selection. In: Proceedings of the 7 th Dialog System Technology Challenge (DSTC7) (2019). https://doi.org/10.1016/j.csl.2020.101072
9. Gašić, M., Jurčiček, F., Thomson, B., Yu, K., Young, S.: On-line policy optimisation of spoken dialogue systems via live interaction with human subjects. In: 2011 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2011, Proceedings. pp. 312–317 (2011). https://doi.org/10.1109/ASRU.2011.6163950
10. Genevay, A., Laroche, R.: Transfer learning for user adaptation in spoken dialogue systems. In: Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS. pp. 975–983 (2016)
11. Jaccard, P.: The distribution of the flora in the alpine zone. New Phytologist **11**(2), 37–50 (1912)
12. Lai, T.L., Robbins, H.: Asymptotically efficient adaptive allocation rules. Advances in Applied Mathematics **6**(1), 4–22 (1985). https://doi.org/10.1016/0196-8858(85)90002-8
13. Levin, E., Pieraccini, R., Eckert, W.: A stochastic model of human-machine interaction for learning dialog strategies. In: IEEE Trans. on Speech and Audio Processing, Vol. 8 (2000)
14. Lin, J.: An exploration of the principles underlying redundancy-based factoid question answering. ACM Trans. Inf. Syst. **25**(2), 6–es (Apr 2007). https://doi.org/10.1145/1229179.1229180, https://doi.org/10.1145/1229179.1229180
15. Liu, B., Yu, T., Lane, I., Mengshoel, O.J.: Customized nonlinear bandits for online response selection in neural conversation models. In: The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18). pp. 5245–5252 (2018)
16. Magarreiro, D., Coheur, L., Melo, F.S.: Using subtitles to deal with out-of-domain interactions. In: SemDial 2014 - DialWatt (2014)
17. McCandless, M., Hatcher, E., Gospodnetic, O.: Lucene in Action, Second Edition: Covers Apache Lucene 3.0. Manning Publications Co., Greenwich, CT, USA (2010)
18. Mendonça, V., Melo, F.S., Coheur, L., Sardinha, A.: A Conversational Agent Powered by Online Learning. vol. 3, pp. 1637–1639. International Foundation for Autonomous Agents and Multiagent Systems, São Paulo, Brazil (2017), http://dl.acm.org/citation.cfm?id=3091282.3091388
19. Mendonça, V., Melo, F.S., Coheur, L., Sardinha, A.: Online learning for conversational agents. In: Lecture Notes in Computer Science (including subseries Lecture

Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10423 LNAI, pp. 739–750 (2017). https://doi.org/10.1007/978-3-319-65340-2_60

20. Oliveira, H.G., Ferreira, J., Santos, J., Fialho, P., Rodrigues, R., Coheur, L., Alves, A.: AIA-BDE: A Corpus of FAQs in Portuguese and their Variations. In: Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020). pp. 5442–5449 (2020)
21. Robbins, H.: Some Aspects of the Sequential Design of Experiments. Bulletin of the American Mathematical Society **58**(5), 527–535 (1952). https://doi.org/10.1090/S0002-9904-1952-09620-8
22. Roller, S., Dinan, E., Goyal, N., Ju, D., Williamson, M., Liu, Y., Xu, J., Ott, M., Shuster, K., Smith, E.M., Boureau, Y.L., Weston, J.: Recipes for building an open-domain chatbot. Tech. rep. (2020), http://arxiv.org/abs/2004.13637
23. Serban, I.V., Sankar, C., Germain, M., Zhang, S., Lin, Z., Subramanian, S., Kim, T., Pieper, M., Chandar, S., Ke, N.R., Rajeswar, S., de Brebisson, A., Sotelo, J.M., Suhubdy, D., Michalski, V., Nguyen, A., Pineau, J., Bengio, Y.: A deep reinforcement learning chatbot. Tech. rep. (2018)
24. Singh, S., Litman, D., Kearns, M., Walker, M.: Optimizing dialogue management with reinforcement learning: Experiments with the njfun system. Journal of Artificial Intelligence Research **16**, 105–133 (2002)
25. Su, P.H., Gašić, M., Mrkšić, N., Rojas Barahona, M.L., Ultes, S., Vandyke, D., Wen, T.H., Young, S.: 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 (Volume 1: Long Papers). pp. 2431–2441. Association for Computational Linguistics (2016). https://doi.org/10.18653/v1/P16-1230, http://aclweb.org/anthology/P16-1230
26. Upadhyay, S., Agarwal, M., Bounneffouf, D., Khazaeni, Y.: A bandit approach to posterior dialog orchestration under a budget. In: 32nd Conference on Neural Information Processing Systems (NeurIPS 2018) (2018)
27. Wang, C.C., Kulkarni, S.R., Poor, H.V.: Bandit problems with side observations. IEEE Transactions on Automatic Control **50**(3), 338–355 (2005). https://doi.org/10.1109/TAC.2005.844079
28. Yu, Z., Xu, Z., Black, A.W., Rudnicky, A.I.: Strategy and Policy Learning for Non-Task-Oriented Conversational Systems. In: Proceedings of the SIGDIAL 2016 Conference. pp. 404–412 (2016)
29. Zhang, Z., Li, J., Zhu, P., Zhao, H., Liu, G.: Modeling Multi-turn Conversation with Deep Utterance Aggregation. In: Proceedings of the 27th International Conference on Computational Linguistics (COLING'18). pp. 3740–3752 (2018), http://arxiv.org/abs/1806.09102