A Conversational Agent Powered by Online Learning (Extended Abstract)

Vânia Mendonça, Francisco S. Melo, Luísa Coheur, Alberto Sardinha
Instituto Superior Técnico / INESC-ID
Av. Prof. Doutor Aníbal Cavaco Silva
Porto Salvo, Portugal
{vania.mendonca, luisa.coheur, jose.alberto.sardinha}@tecnico.ulisboa.pt, fmelo@inesc-id.pt

ABSTRACT

In this work, we improve the performance of a dialogue engine, Say Something Smart, using online learning. Given a request by a user, this engine selects an answer from a corpus of movie subtitles, weighting the quality of each candidate answer according to several criteria and selecting the one that is chosen by the most representative criteria. We contribute with an online approach, using sequential learning, that adjusts the weights of the different criteria using a reference corpus of actual dialogues as input to simulate user feedback. This approach effectively allowed Say Something Smart to improve its performance at each interaction, as shown in an experiment performed in a test corpus.

CCS Concepts

•**Computing methodologies** → *Intelligent agents; Discourse, dialogue and pragmatics;*

Keywords

Conversational agents; online learning; Exponentially Weighted Average Forecaster

1. INTRODUCTION

Conversational agents are becoming ubiquitous in our lives, either as personal assistants in mobile devices (e.g., Apple's Siri, Microsoft's Cortana), or as guides in public places (e.g., Edgar Smith, at the palace of Monserrate [5], Ada and Grace, at the Museum of Science in Boston [12]). Most of these agents are designed to operate within a well-defined domain. However, users often pose out-of-domain requests to test the extent to which the agent can keep a conversation. Several agents handle such requests by fetching answers from a pre-defined set or by explaining their inability to answer as a result of certain "human" characteristics of the agent.

Recent research has explored the possibility of addressing such requests without resorting to hand-crafted responses, considering that it is impossible to address all potential requests from users. Such systems resort to large collections of data from which they extract potential answers to those requests. This is the case of the agent Filipe, which provides answers based on interactions extracted from movie subtitles [1, 8]. Filipe uses Say Something Smart (SSS), an engine that compares a given user request with the interactions present in the corpus, scoring them according to a set of configurable weighted criteria, and selects the most voted answer, according to those criteria. The weights assigned to each criteria in the work of Magarreiro et al [8] were empirically chosen; our hypothesis is that superior performance could be attained by learning those weights based on feedback. Several works have used learning strategies that incorporate feedback, achieving promising results [6, 10, 9, 11] (see Cuayahuitl & Dethlefs [3] for a review).

Therefore, the key contribution of this paper is a novel application of an online learning approach tailored for an open domain conversation agent. We use sequential learning within SSS to learn the weights from a corpus of dialogues (simulating user feedback). We compare the performance of the learnt weights to those reported in the work of Magarreiro et al [8], in an experiment where user feedback is simulated from corpora.

2. SAY SOMETHING SMART

Say Something Smart (SSS) is the dialogue engine behind the agent Filipe [1, 8]. Given a user request, SSS looks up for a set of answer candidates in a corpus of interactions and returns the best answer according to a combination of several configurable criteria. The corpus in use, Subtle, is composed of pairs of consecutive subtitles extracted from OpenSubtitles[1], where the first element of the pair is the *trigger*, and the second is the *answer*.

For each request u, a set of up to N candidate interactions $C = \{c_1, \ldots, c_N\}$ is retrieved, where each interaction c_n is a trigger-answer pair, (T_n, A_n), and N is a configurable value. SSS then has each criterion comparing every interaction $c_n \in C$ to the user input u and scoring them. The total score of each interaction $c_n \in C$ is given by:

$$score(c_n) = \sum_{k=1}^{K} w_k M_k(C, T_n, A_n, u), \tag{1}$$

where w_k is the weight associated with criterion M_k. Then, given the interaction

$$c^* = \operatorname*{argmax}_{c_n \in C} score(c_n),$$

SSS outputs the answer associated with c^* as the reply to the user input u [8].

[1] `http://www.opensubtitles.org/`

Appears in: *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

3. SEQUENTIAL LEARNING APPLIED TO SAY SOMETHING SMART

The performance of SSS depends critically on two factors, namely (i) the ability of criteria $M_1, \ldots, M_K$ to identify plausible interactions given a user input; and (ii) a correct setting of the weights $w_1, \ldots, w_K$ that balances the contribution of the different criteria in the selection of the best interaction. As reported in the work of Magarreiro et al [8], the weights in SSS were "empirically set". We instead propose that these weights are *learnt* from feedback provided by human users, by accommodating this feedback in the process of selecting the weights assigned to each of the criteria used to assess the quality of different answers.

This imposes two important requirements: (i) the algorithm should learn *incrementally* from successive feedback, allowing the performance of the system to immediately incorporate each piece of feedback provided by users; (ii) the learning algorithm used should be *fast* at incorporating user feedback, since user interactions are potentially expensive. As such, we adopt an online approach, choosing a standard sequential learning algorithm known as Exponentially Weighted Average Forecaster (EWAF), a generalization of the Weighted Majority algorithm of Littlestone & Warmuth [7]. EWAF precisely addresses these requirements: it has well-established performance guarantees, which include a bound on how fast it converges [2].

In a first approach to validate the proposed learning approach, we used a reference corpus to simulate the user feedback: a set of pairs $(trigger, answer)$ from Cornell Movie-Dialogs (CMD) corpus [4], which contains over $80,000$ conversations corresponding to actual movie scenes (something that is not guaranteed in the case of Subtle interactions).

The learning process proceeds as follows. At each step t, a "user interaction" $u(t) = (T_{u(t)}, A_{u(t)})$ is selected from the reference corpus and the trigger $T_{u(t)}$ is presented to SSS as being a user request. SSS retrieves a set $C(t) = \{c_1(t), \ldots, c_N(t)\}$ of candidate interactions and each interaction $c_n(t) \in C(t)$ is scored according to Equation 1. The interaction $c^* = (T^*, A^*)$ with maximum score is computed according to each criterion M_k. The criteria are evaluated considering their choices for the best interactions with a value of "user feedback" $r_k(t)$, and the weights $w_1, \ldots, w_K$ are updated as a function of $r_k(t)$. In other words, we use the reference corpus as user feedback to "train" SSS.

4. EVALUATION

To evaluate our contribution, we defined the following research question: *Can iteratively learnt weights outperform the handcrafted weights reported as best by Magarreiro et al [8]?* To address this question, we devised an experiment comprising two phases. In the first phase, we learn sets of weights w_k using different configurations of the algorithm's meta-parameters and assess the performance of those sets of weights in order to choose a meta-parameter configuration. To obtain the sets of weights, we performed 6 runs per combination of meta-parameters, each using a different subset of CMD as input (each subset has 1000 interactions). Then, we evaluated the performance of the different sets of weights by running SSS using a subset of CMD, containing 2000 interactions, as both the input and the source of subtitles (instead of using Subtle).

For each set, we computed the accuracy of the system, i.e., the percentage of iterations in which SSS was able to choose the candidate answer that matched the input reference answer. In the second phase, we compared the performance of the weights learnt using the configuration chosen in the first phase against the ones reported as best by Magarreiro et al [8]. We had SSS learning using that combination over an input set composed of the 6 training input subsets, shuffled. At each 500 iterations, we "froze" the weights obtained and run SSS with them to assess their performance: we used a (different) subset of CMD, containing 2000 interactions, as both the input and the source of subtitles, and then we computed the accuracy, similarly to the first phase. Finally, we compared the evolution of the accuracy achieved by the weights learnt at each point against the accuracy achieved by the weights reported by Magarreiro et al [8].

We present the evolution of the accuracy obtained with the weights learnt at each 500 iterations in Figure 1 (blue line). Each diamond-shaped point represents the accuracy obtained by the set of weights learned at a given iteration. In the first 500 iterations, both the learnt weights and the ones from Magarreiro et al [8] are tied with an accuracy of 87.15%, and from then on, the learnt weights improve the system's accuracy to 95.2% and stabilize.

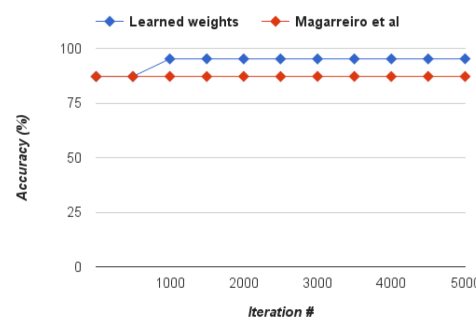

Figure 1: Comparison between the performance of the weights reported as best by Magarreiro et al [8] and the weights learnt by our learning module.

5. DISCUSSION

The results reported above suggest that learning weights using an online approach instead of using fixed weights indeed improves the performance of SSS. However, one might argue that the gain from learning is not significant, as it stabilized very early. In fact, in this situation the fixed weights selected by Magarreiro et al [8] were already capable of achieving ah high accuracy value, which left little room for improvements, but, taking this experiment to different settings, it could be the case that the initial weights would be far from optimal, in which case learning the weights could attain a more significant improvement.

Acknowledgments

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013, and under project CMUP-ERI/HCI-/0051/2013. Vânia Mendonça is funded by an FCT scholarship with reference SFRH/BD/121443/2016.

REFERENCES

[1] D. Ameixa, L. Coheur, P. Fialho, and P. Quaresma. Luke, I am your father: Dealing with out-of-domain requests by using movies subtitles. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8637 LNAI:13–21, 2014.

[2] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning and Games*. Cambridge University Press, 2006.

[3] H. Cuayáhuitl and N. Dethlefs. Dialogue systems using online learning: Beyond empirical methods. In *NAACL-HLT Workshop on Future Directions and Needs in the Spoken Dialog Community: Tools and Data*, SDCTD '12, pages 7–8, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

[4] C. Danescu-Niculescu-Mizil and L. Lee. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, ACL 2011*, 2011.

[5] P. Fialho, L. Coheur, S. Curto, P. Cláudio, Â. Costa, A. Abad, H. Meinedo, and I. Trancoso. Meet edgar, a tutoring agent at monserrate. In *ACL (Conference System Demonstrations)*, pages 61–66. The Association for Computer Linguistics, 2013.

[6] E. Levin, R. Pieraccini, and W. Eckert. A stochastic model of human-machine interaction for learning dialog strategies. In *IEEE Trans. on Speech and Audio Processing, Vol. 8*, 2000.

[7] N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Inf. Comput.*, 108(2):212–261, Feb. 1994.

[8] D. Magarreiro, L. Coheur, and F. S. Melo. Using subtitles to deal with out-of-domain interactions. In *SemDial 2014 - DialWatt*, 2014.

[9] O. Pietquin, M. Geist, and S. Chandramohan. Sample Efficient On-line Learning of Optimal Dialogue Policies with Kalman Temporal Differences. In *International Joint Conference on Artificial Intelligence (IJCAI 2011)*, pages 1878–1883, Barcelona, Spain, July 2011.

[10] S. Singh, D. Litman, M. Kearns, and M. Walker. Optimizing dialogue management with reinforcement learning: Experiments with the njfun system. *Journal of Artificial Intelligence Research*, 16:105–133, 2002.

[11] P.-H. Su, M. Gasic, N. Mrkšić, M. 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 (Volume 1: Long Papers)*, pages 2431–2441. Association for Computational Linguistics, 2016.

[12] D. Traum, P. Aggarwal, R. Artstein, S. Foutz, J. Gerten, A. Katsamanis, A. Leuski, D. Noren, and W. Swartout. *Ada and Grace: Direct Interaction with Museum Visitors*, pages 245–251. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.