An Online Platform for crowd-Sourcing Data from Interactions with Chatbots

WOCHAT SHARED-TASK REPORT

Luis Fernando D'Haro[1], Lue Lin[2]

[1]Institute for Infocomm Research, [2]Ngee Ann Polytechnic, Singapore
luisdhe@i2r.a-star.edu.sg, llue97@hotmail.com

Abstract. Chatbots are increasingly gaining interest given their potential to provide quick answers and entertainment to final users on different channels (web-site, social networks, or messaging apps). Unfortunately, creating data-driven systems is not easy as they require huge amount of annotated human-chatbot dialogs which are not currently available. In this paper, we describe a crowd-sourcing online platform[1] to record human-chatbot interactions that can be evaluated on-demand after each turn is given or offline by registered volunteers, with the goal of generating a publicly anonymized database for the research community.

Keywords. Chatbots, data collection, annotation, online service.

1 Introduction

By definition, a chatbot is a computer program that responds to natural language text and/or to voice inputs in a human like manner. Their main advantage is that they can perform many tasks given specific commands while they give the impression that they can understand, talk and entertain users. Chatbots have been around for decades, but they have gained a lot of interest in the last years because of challenges like the Loebner prize[2] where the main goal is to create human-like chatbots, and because companies can use them to quickly provide corporate information on their websites like Ask Anna from Ikea [1], on social networks like Facebook Messenger, on mobile apps like Apple Siri [2], Cortana, and recently messaging apps like Xiaoice [10], a tendency that is expected to increase when we consider that in average a teenager spend around 2.4 hours daily texting or emailing [6].

Although a basic chatbot can be easily built by modifying available knowledge-based rules[3] and using platforms like pandorabot[4,] the reality is that their maintenance is a time consuming task, and most data-driven approaches cannot be used since the

[1] http://www.teachabot.com:8000/main

[2] http://www.loebner.net/Prizef/loebner-prize.html

[3] http://www.alicebot.org/

[4] http://www.pandorabots.com/

number of available conversational databases is very small and interactions collected by existing chatbots are undisclosed due to confidentiality.

In this paper, we introduce, up to the best of our knowledge, the first crowd source initiative to collect, annotate and publicly release a database of human-machine chat interactions. For this, we have deployed an online website where human-chatbot dialogs can be collaboratively collected and evaluated. The platform also allow: a) downloading existing plain and annotated sessions, and b) the easy integration of existing chatbots by using websockets or REST calls using few simple JSON messages[5].

This paper is organized as follows: section 2 shows the architecture and main capabilities of the platform and section 3 presents the conclusions and future work.

2 Platform description and capabilities

The main goals of the proposed platform are: a) To provide an online platform where registered participants will generate and annotate human-machine dialogs, b) To provide an easy to use online platform for evaluating chatbots, and c) To allow registered participants to download existing recorded and annotated interactions for further research and experimentation.

2.1 Architecture and Normal usage

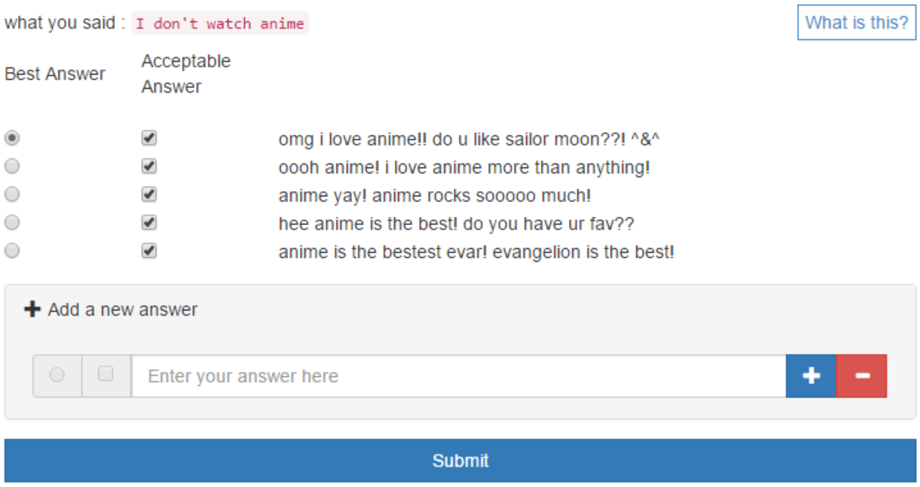
The platform consists of three main components: 1) a scalable and multi-thread server implemented in Tornado that allows the connections between users and chatbots by using websockets or REST calls with Webhooks, which records all interactions on a NonSQL database (MongoDB), as well as control all connections and messages between chatbots and clients, 2) a mobile-first responsive web-site implemented using Bootstrap and Javascript that acts as a front-end for connecting users and chatbots, and 3) the chatbots that connect to the server to interchange messages with the users using JSON messages.

The main interface shows a list of available chatbots that users can select to interact with. However, users need first to login providing a username and email (information that is only used to keep track of the history of chat sessions). Once logged in, users can perform offline annotations.

During a typical human-chatbot interaction, the system creates a chatroom where the user and the chatbot engage in a one-by-one series of turns until the user leaves the room or logout from the platform. During the interactions, the user has the possibility of performing an online annotation process by activating an evaluation mode where, for each user’s turn, the system shows a N-best list of suitable answers that the chatbot can generate (see figure 1). In this case, the user can decide which answers are valid ones and which one is the best possible answer. Optionally, users can add new answers if they want. Once the form is processed, the best answer is displayed to the user and the chatbot receives this information to keep track of the dialog. When the

[5] http://workshop.colips.org/wochat/webchat_api.html

user finishes chatting, the user is requested to evaluate the chatbot and provide general comments (information that is also sent to the chatbot for its own record).



what you said : I don't watch anime

Best Answer	Acceptable Answer	
●	☑	omg i love anime!! do u like sailor moon??! ^&^
○	☑	oooh anime! i love anime more than anything!
○	☑	anime yay! anime rocks sooooo much!
○	☑	hee anime is the best! do you have ur fav??
○	☑	anime is the bestest evar! evangelion is the best!

Fig. 1. Form to evaluate N-best lists of answers allowing users to add new ones.

2.2 Baseline chatbots

In order to allow developers to quickly integrate their chatbots into the platform, we implemented 4 basic chatbots based on NLTK[6] (i.e. Eliza, Zen master, Iesha and Rudeman) whose code is available for downloading from the main interface [4]; also, we include examples of how to connect them to the server platform by using websockets or WebHooks (i.e. bidirectional REST calls). In addition, users are allowed to download previous chat sessions in different formats such as XML or JSON).

2.3 Annotation and evaluation

In the literature we can find several metrics to evaluate and compare chatbots [8, 3, 7] like adequacy, polarity, number of turns, total time of interaction, semantic similarity and agreement between human and chatbot turns, subjective evaluations, etc. The main purpose of these metrics is to detect failures in the understanding and generation of the chatbot answers, and to identify good and bad users' questions and answers.

Although there is a long road to have intelligent systems that can understand, communicate and learn from users as proposed in [5], the platform is targeted to facilitate the process of annotating the generated corpus including information about the quality of the interactions in terms of adequacy, polarity, presence of offensive or swearing language, duration of each turn, total number of turns, etc. Part of this information is automatically collected from the user after each turn or at the end of the chatting session (see figure 1), or by detailed annotations manually generated by vol-

[6] http://www.nltk.org/api/nltk.chat.html

unteers that go sequentially through a list dialogs and are requested to evaluate every turn only based on the current history of the dialog without considering the following turns. Since the turns and participants in the chat room are anonymized, it is possible to evaluate whether annotators can detect human from machine turns, as well as to use the final annotations to train classifiers and re-rankers for generating answers.

3 Conclusions and future work

In this paper we have described an online platform that allows the recording, evaluation and annotation of human-chatbots dialogs with the purpose of generating a publicly available corpus for researchers. The platform allows users to compare different chatbot answers for quickly improving generative models. This paper is intended to encourage researchers to participate in the process of generating content, providing chatbots or annotating the dialogs, as well as to use the platform to support existent workshops and shared tasks related to chatbot evaluation. As future work, we are plan to extend the list of available chatbots by incorporating DNN-based chatbots [9].

4 References

1. Artificial-Solutions: Case Study: Artificial Solutions Enables IKEA to Self-Serve Ask Anna (Feb 2016), http://marketing.artificial-solutions.com/rs/artificialsolutions/images/CS_Anna.pdf.
2. Bellegarda, J.R.: Spoken language understanding for natural interaction: The siri experience. In: Natural Interaction with Robots, Knowbots and Smartphones, pp. 3–14. Springer (2014)
3. Hung, V., Elvir, M., Gonzalez, A., DeMara, R.: Towards a method for evaluating naturalness in conversational dialog systems. In: Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on. pp. 1236–1241. IEEE (2009)
4. Li, L., D'Haro, L.F.: Regular Expression Based Agents for Online Collection of Human-Chatbot Interactions. Proceedings Workshop IVA 2016
5. Mikolov, T., Joulin, A., Baroni, M.: A roadmap towards machine intelligence. arXiv preprint arXiv:1511.08130 (2015)
6. Roberts, J., Yaya, L., Manolis, C.: The invisible addiction: Cell-phone activities and addiction among male and female college students. Journal of Behavioral Addictions 3(4), 254–265 (Dec 2014), http://www.akademiai.com/doi/abs/10.1556/JBA.3.2014.015
7. Serban, I.V., Lowe, R., Charlin, L., Pineau, J.: A Survey of Available Corpora for Building Data-Driven Dialogue Systems. arXiv preprint arXiv:1512.05742 (2015)
8. Shah Huma, Warwick Kevin, Vallverd´u Jordi, Wu Defeng: Can machines talk? Comparison of Eliza with modern dialogue systems. Computers in Human Behavior 58, 278–295 (may 2016), http://www.sciencedirect.com/science/article/pii/S0747563216300048
9. Vinyals, O., Le, Q.: A neural conversational model. arXiv preprintarXiv:1506.05869 (2015)
10. Wang, Y.: Your Next New Best Friend Might Be a Robot: Meet Xiaoice. She's empathic, caring, and always available—just not human. Nautilus (Feb 2016), http://nautil.us/issue/33/attraction/your-next-new-best-friend-might-be-a-robot.