CHATBOT using Deep Learning (Seq2Seq Models)

Gaurav Rajpoot, Arul Srivastava, Dilip Kumar, Monica Sehrawat

Abstract— **Many conversational agents (CAs) are developed to answer users' questions in a specialized domain. In everyday use of CAs, user experience may extend beyond satisfying information needs to the enjoyment of conversations with CAs, some of which represent playful interactions. By studying a field deployment of a Human Resource Chabot, we report on users' interest areas in conversational interactions to inform the development of CAs. Through the lens of statistical modeling, we also highlight rich signals in conversational interactions for inferring user satisfaction with the instrumental usage and playful interactions with the agent. These signals can be utilized to develop agents that adapt functionality and interaction styles. By contrasting these signals, we shed light on the varying functions of conversational interactions. We discuss design implications for CAs, and directions for developing adaptive agents based on users' conversational behaviors.**

Index Terms— **Conversational agent; Chabot; dialog system; human-agent interaction; playful; adaption; user modeling.**

I. INTRODUCTION

Highlight There is a growing excitement around conversational agents (CAs) or "chatbots". From tech giants' core products such as Apple Siri, Amazon Alexa, IBM Watson, to numerous startup companies, many are compelled by the idea of advances in Many conversational agents (CAs) are developed to answer users' questions in a specialized domain. In everyday use of CAs, user experience may extend beyond satisfying information needs to the enjoyment of conversations with CAs, some of which represent playful interactions. By studying a field deployment of a Human Resource chatbot, we report on users' interest areas in artificial intelligence combined with a natural form of interactions. However, before this wave of marketing hype, research on CAs has come a long way in the past half century, but also saw several unfortunate failures in public reception. Two points of criticism have been frequently raised for studies of CAs. One is a lack of understanding on real-life user experience and attention to the gap between user interactions in the lab and those in the wild .Another point is the focus on narrowly constrained agent initiated conversations for the task domains, which provides little information about user interests in conversing with CAs for future system development [38]. Although many recent popular CAs, often in the form of an intelligent personal assistant, provide free-form text input interfaces that invite users to "ask me

**Gaurav Rajpoot,** B. Tech Scholar (CSE), ABES Institute of Technology, Ghaziabad

Arul Srivastava, B. Tech Scholar (CSE), ABES Institute of Technology, Ghaziabad

Dilip Kumar, B. Tech Scholar (CSE), ABES Institute of Technology, Ghaziabad

Monica Sehrawat, Assistant Prof. Department of Computer Science ABES Institute of Technology, Ghaziabad

anything", there is surprisingly limited empirical account of how users converse with these agents in the wild. This poses a challenge as the development of CAs, at least in the near term, relies heavily on the anticipation of what users may say to the agent. To fill these gaps, we study a field deployment of a question and-answer (QA) CA. Specifically, a Human Resource (HR) chatbot provided company-related information assistance to 377 new employees for 6 weeks. Although the CA functions as a QA system, the focus of this paper is on users' conversational interactions, or social dialogues, with the agent (36% of interaction logs). Such interactions are often abundant as CAs naturally elicit social behaviors with a human role metaphor. Meanwhile, there is a tradition of separating communicative and task-oriented interactions in developing CAs, and considering the former to be more unbounded and thus challenging to anticipate, but have the advantage of higher generalizability across domains[38].The generalizability motivated some to build domain-independent conversation architectures to accelerate the development of CAs. Also underscoring the necessity of studying conversational interactions is the rich signals they may carry for inferring user status. This has been a long standing interest in the related embodied conversational agents (ECA) and human-robot interaction (HRI) communities, as foundational work to build adaptive agents that can attend to user needs. For example, by inferring a decline in user engagement, an agent can immediately employ strategies to re-engage the user .

To infer such an internal user status, agents rely on recognizing signals in users' behavioral manifestation. For example, gaze fixation and attentive feedback ("un-huh") are signals of engagement. These association rules are an integral part of the computational models underlying adaptive agents. However, most existing work drew on observations from human-human communications, and aimed to infer human concepts of inter-personal status such as rapport and trust . In the context of un-embodied QA agents, the system is not intended to serve full conversations, and some users may simply reject to anthropomorphize QA agents such as Siri. It is arguable whether these behavioral signals knownfromhumanconversationsstillhold. Itisalsoarguable whether human-like inter-personal status should be the primary dimension that a QA agent is concerned with. Deep neural networks contain multiple non-linear hidden layers and this makes them very expressive models that can learn very complicated relationships between their inputs and outputs. With limited training data, however, many of these complicated relationships will be the result of sampling noise, so they will exist in the training set but not in real test data even if it is drawn from the same distribution. This leads to overfitting and many methods have been developed for reducing it. These include stopping the training as soon as

performance on a validation set starts to get worse, introducing weight penalties of various kinds such as L1 and L2 regularization and soft weight sharing (Nowlan and Hinton, 1992).

With unlimited computation, the best way to "regularize" a fixed-sized model is to average the predictions of all possible settings of the parameters, weighting each setting by
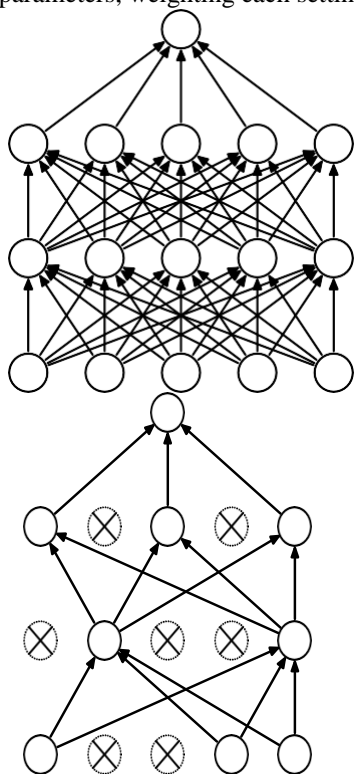
Its posterior probability given the training data. This can sometimes be approximated quite well for simple or small models (Xiongetal.,2011; Salakhutdinov and Mnih, 2008), but we would like to approach the performance of the Bayesian gold standard using considerably less computation. We propose to do this by approximating an equally weighted geometric mean of the predictions of an exponential number of learned models that share parameters.

Conversational Interactions with CAs:

Development of CAs can be dated back to the 1950s with prominent examples such as ELIZA . Within the HCI field, research largely focused on embodied CAs. Anthropomorphism is emphasized in multiple modalities to regulate human-computer interactions in a familiar way and to manifest social intelligence such as trust worthiness . Recently, the term "chatbot" is used to refer to CAs that employ primarily text-based or speech-based input without embodied modalities. This type of CA has become mainstream products. Some argue that for these systems, anthropomorphism is no longer a principal goal, and the single modality directs more attention to task performances, especially since many of these CAs are core components of utility applications. Despite the anti-anthropomorphism argument, the interaction is still based on the metaphor of human conversations, which is a complex machinery in its own right. but can also diverge from human conversations in

many ways .

This highlights the necessity of studying patterns of conversational interactions with CAs, which can be considered as user utterances in performing communicative and social functions instead of task-oriented functions (e.g.,QA query). Early systems often adopted agent controlled conversations to avoid the daunting challenge of handling unbounded conversations initiated by users (e.g. [11]). But this approach is inadequate for realistic conversational capabilities, and obsolete for QA agents that provide information assistance through free-form text input. At the present time, whether using a rule-based system or advanced technologies such as discourse planners, the development of CAs relies heavily on the anticipation of what users may say to the agent. Ignoring common patterns may result in the absence of necessary system knowledge, and thus repeatedly frustrating "sorry I didn't get it" responses. To overcome the problem, development of CAs has to follow a laborious iterative process to bootstrap from user data .

II. SYSTEM DESCRIPTION
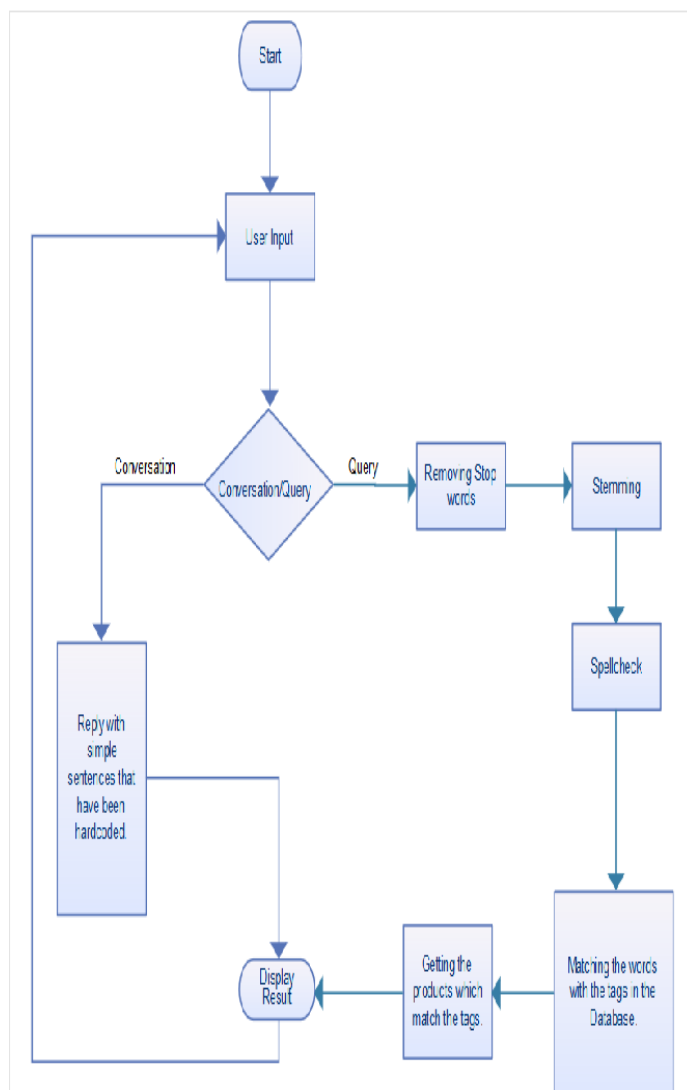
Natural language Classifiers (NLCs) and Performance:
For the analysis, we utilized the natural language classifiers (NLCs) of the system to provide a characterization of users' conversational input. Here we briefly discuss the NLCs to pro- vide background for our methodology. The technical

details of the NLCs are beyond the scope of this paper. Many current CA technologies rely on NLCs to classify a user input (e.g., "hello") into a higher-level category of intent (e.g., "GREET- ING") known to the system in order to retrieve answers. Chip adopted a multi-level NLC model by independently training two levels of NLCs, each as a multi-class classifier (i.e., an input is classified to be the intent class with the highest confidence score). NLC1 contains higher-level categories of intent, each has several matching NLC2 sub-categories. For example, when a user asks "tell me about health benefits", it will be classified as "BENEFITS" by NLC1, and "health benefits" by NLC2, independently. "Health benefits" is a sub-intent known to match "BENEFITS", which also has other matching sub-intents such as "dental plan" and "employee discount". Each NLC2 class is linked to a curated answer, sometimes with variations to be randomly retrieved. For example, in this case, it will output the curated answer linked to "health benefits" to answer the user question.

MOST IMPORTANT TOOLS USED:
Defined in file: TensorFlow/python/ops/gen_array_ops.py.
- tf.fill
- tf.strided_slice
- tf.concat
- tf.nn.rnn_cell.BasicLSTMCell
- tf.nn.embedding_lookup

III. LIMITATIONS

As with any automatic methods to characterize large quantity of texts, the conversational labels we obtained were not without noise. The schema we used might not capture rarer cases of conversational interactions. However, our goal is not to establish a formal taxonomy but to contribute empirical insights by quantifying available types of conversational acts. Lastly, we acknowledge that some observations may be specific to the workplace context and user sample of the study, as young professionals may be more inclined for playful interactions. We do not claim the generalization of specific statistics but focus on the patterns they represent.

IV. CONCLUSION

By studying log data from a field deployment of a question-and-answer conversational agent, we characterize the rich forms of conversational interactions users had with theagent. Themainareasofconversationsincludefeedback-giving,play-fulchit-chat,systeminquiry,andhabitualcommunicativeutter-ances. Through the lens of statistical modeling, we highlight therichsignalsinconversationalinteractionsforinferringuser satisfaction, which can be utilized to develop agents that can adapt algorithmic performances and interaction styles. The resultsalsoprovidenuancedunderstandingontheunderlying functions of conversational behaviors with QA agents and theirdeviationsfromhumanconversations. Ourfindingsmay inform designs of CAs and contribute to the emerging fields of conversational UX, conversational IR and adaptive agents.

REFERENCES

[1] John Langshaw Austin. 1975. *How to do thingswith words*. Oxford universitypress.
[2] Niels Bernsen and Laila Dybkjær.2004. Domain-oriented conversation with HC Andersen.*Affective Dialogue Systems* (2004), 142–153.
[3] Yulong Bian, ChengleiYang, Dongdong Guan, Sa Xiao, Fengqiang Gao, Chia Shen, and Xiangxu Meng. 2016. Effects of Pedagogical Agent's Personality and Emotional Feedback Strategy on Chinese Students' Learning Experiences and Performance: A Study Based onVirtualTaiChiTrainingStudio.In*Proceedingsofthe 2016 CHI Conference on Human Factors in Computing Systems*. ACM,433–444.
[4] TimothyBickmore,LauraPfeifer,andDanielSchulman. 2011. Relational agents improve engagement and learning in science museum visitors. In *International Workshop on Intelligent Virtual Agents*. Springer,55–67.
[5] Timothy W Bickmore, Laura M Pfeifer, and Brian W Jack. 2009. Taking the time to care: empowering low health literacy hospital patients with virtual nurseagents. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM,1265–1274.
[6] DanBohusandAlexanderIRudnicky.2003.RavenClaw: Dialog management using hierarchical task decomposition and an expectation agenda.(2003).
[7] SusanEBrennan.1990.Conversationasdirect manipulation: An iconoclastic view.(1990).
[8] Hendrik Buschmeier and Stefan Kopp. 2011.Towards conversational agents that attend to and adapt to communicative user feedback. In *Intelligent Virtual Agents*. Springer,169–182.
[9] John M Carroll and John C Thomas. 1988. Fun.*ACM SIGCHI Bulletin* 19, 3 (1988),21–24.
[10] JustineCassell.2001.Embodiedconversationalagents: representation and intelligence in user interfaces. *AI magazine* 22, 4 (2001),67.
[11] JustineCassellandTimothyBickmore.2003.Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User modelingand user-adapted interaction* 13, 1 (2003), 89–132.
[12] Praveen Chandar, YasamanKhazaeni, MatthewDavis, Michael Muller, Marco Crasso, Q Vera Liao, N Sadat Shami, and Werner Geyer. 2017. Leveraging Conversational Systems to Assists New Hires During Onboarding. In *IFIP Conference onHuman-Computer Interaction*. Springer,381–391.
[13] Mark Coeckelbergh. 2011. You, robot: on the linguistic constructionofartificialothers.*AI&society*26,1(2011), 61–69.
[14] Mark G Core and James Allen. 1997. Codingdialogs with the DAMSL annotation scheme. In *AAAI fall symposium on communicative action in humans and machines*, Vol. 56. Boston,MA.
[15] Doris M Dehn and Susanne Van Mulken. 2000.The impact of animated interface agents: a review of empirical research. *International journalof human-computer studies* 52, 1 (2000), 1–22.
[16] Nicholas Epley, Adam Waytz, and John T Cacioppo. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychological review* 114, 4(2007), 864.
[17] Dan Fletcher. 2010. The 50 Worst Inventions:Microsoft Bob. TIME. (27 May2010).
[18] Eric Gilbert. 2012. Phrases that signal workplace hierarchy.In*ProceedingsoftheACM2012conferenceon Computer Supported Cooperative Work*. ACM, 1037–1046.
[19] Joakim Gustafson and Linda Bell. 2000. Speech technologyontrial:ExperiencesfromtheAugustsystem. *Natural Language Engineering* 6, 3-4 (2000),273–286.
[20] John Heritage and John Maxwell Atkinson. 1984. *Structures of social action: Studies inconversation analysis*. Cambridge UniversityPress.
[21] Jiepu Jiang, Ahmed Hassan Awadallah, RosieJones, UmutOzertem, ImedZitouni,Ranjitha Gurunath Kulkarni, and Omar Zia Khan. 2015. Automatic online evaluation of intelligent assistants. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 506–516.
[22] Diane Kelly and Jaime Teevan. 2003. Implicitfeedback for inferring user preference: a bibliography. In *ACM SIGIR Forum*, Vol. 37. ACM,18–28.
[23] AlfredKobsaandWolfgangWahlster.1989.*Usermodels in dialog systems*.Springer.
[24] Stefan Kopp, Lars Gesellensetter, Nicole C Krämer,and IpkeWachsmuth. 2005. A conversational agent as museum guide–design and evaluation of a real-world application. In *International Workshop on Intelligent Virtual Agents*. Springer,329–343.
[25] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2010. Receptionist or information kiosk: How do peopletalk with a robot?. In *Proceedings of*

the 2010 ACM conference on Computer supported cooperative work. ACM,31–40.

[26] Namseok Lee, Hochul Shin, and S Shyam Sundar.2011. Utilitarianvs.hedonicrobots:roleofparasocialtendency and anthropomorphism in shaping user attitudes. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on.* IEEE,183–184.

[27] Q Vera Liao, Matthew Davis, Werner Geyer, Michael Muller, and N Sadat Shami. 2016. What Can You Do?: Studying Social-Agent Orientation and Agent Proactive Interactions with an Agent for Employees. In *Proceedingsofthe2016ACMConferenceonDesigning Interactive Systems.* ACM,264–275.

[28] Ewa Luger and Abigail Sellen. 2016. Like Having a Really Bad PA: The Gulf between User Expectation and ExperienceofConversationalAgents.In*Proceedingsof the 2016 CHI Conference on Human Factors in Computing Systems.* ACM,5286–5297.

[29] ChrisMatyszczyk.2012.Apple'sSiriwrong38percent of the time in test. CNET. (30 June2012).

[30] Tanushree Mitra and Eric Gilbert. 2014. The language that gets people to give: Phrases that predict successon kickstarter. In *Proceedings of the 17th ACMconference on Computer supported cooperative work & social computing.* ACM,49–61.

[31] Robert J. Moore, Rafah A. Hosn, and Ashima Arora. 2016. The Machinery of Natural Conversation andthe Design of Conversational Machines. In *American Sociological Association annualmeeting.*

[32] Amy Ogan, Samantha Finkelstein, Elijah Mayfield, ClaudiaD'Adamo,NoboruMatsuda,andJustineCassell. 2012a.Ohdearstacy!:socialinteraction,elaboration,and learning with teachable agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM,39–48.

[33] Amy Ogan, Samantha L Finkelstein, Erin Walker, Ryan Carlson, and Justine Cassell. 2012b. Rudeness and Rapport: InsultsandLearningGainsinPeerTutoring.