

Toward Narrative-Based Conversational Interfaces

Ethel Chua Joy Ong

Center for Language Technologies, De La Salle University
2401 Taft Avenue, Malate, Manila, 1004 Philippines
ethel.ong@dlsu.edu.ph

Abstract

Advances in language technology research led to the rise in popularity of conversational interfaces that allow human users to carry on natural-like conversations with the software agent during task performance. Children comprise a sector of society who can benefit from interacting with these agents, particularly to enhance their language acquisition, literacy skills and emotional development. Stories dominate the everyday conversations of children. They learn about themselves and their environment through sharing of stories. In this paper, we describe our conversational storytelling agent that can support children in constructing a story, reflecting on story events, and expressing their emotions.

Keywords: conversational interfaces, conversational agents, collaborative storytelling, joint story reading, dialogue

1. Introduction

The rising popularity of conversational interfaces, notably voice assistants such as Apple's Siri, Amazon's Alexa, Google Assistant and Microsoft Cortana, is geared toward setting up a more natural interface wherein the human user and the computer are able to interact meaningfully when carrying out certain tasks. Despite the achievement of these technologies in terms of the number of available *skills* and *actions* (Dale, 2019), their capabilities are still limited to specific goal-oriented functions, such as searching for information to answer user queries, recommending a route, reminding users about upcoming events, playing the user's favorite music, and scheduling appointments for the user.

Fay (2014) believed that these voice assistants cannot yet engage their human users in everyday conversations akin to natural human discourse. A key missing ingredient, he noted, is the conversational agent's capacity to understand and generate stories, which is inherently embedded in the human thought process. Everyday human conversations are marked by features of storytelling, becoming free-flowing exchange of information on one's life events and experiences. Narrative intelligence affords people the ability to organize a given series of events into stories (Blair and Meyer, 1997).

As computers become more ubiquitously involved in our day-to-day activities, software agents used in conversational interfaces should embody some form of narrative intelligence to enrich their conversations with human users. Storytelling strategies can be used to elicit further details regarding a user's needs, and to analyze user requests in the context of his/her environment.

Recent advances in language technologies can support the development of conversational agents exhibiting storytelling abilities, with applications in collaborative child-agent storytelling and joint story reading, to encourage children to express their ideas while developing their linguistic and literacy skills. The agent can function as a facilitator, a tutor, or a learning peer, thus providing opportunities for itself and the human user to augment each other's storytelling abilities. In this paper, we describe our conversational storytelling agents, discuss the challenges we encountered during their development, and recommend further work to extend their capabilities.

2. Related Work

Conversational agents are virtual agents designed to mimic human capabilities and attributes, and to collaborate with human users to facilitate thinking and decision-making during the performance of specific tasks. They utilize various means of communication, from written text to voice inputs, in order to "engage in dialogues, and negotiate and coordinate the transfer of information" (Coen, 1995).

2.1 Roles of Conversational Agents

Conversational agents are currently predominantly found in commercial applications as personal voice assistants and service-performing agents. Designed primarily to provide customer support, they can assist in product search, answer customer queries, and perform simple well-defined tasks requested by the user.

Since conversational agents' roots can be traced to intelligent software agents, they can also take on various roles depending on the types of interaction expected from them, as shown in previous research. These include counselor, critic, facilitator, tutor, and peer.

As a counselor, the agent promotes certain behaviors and practices, such as in healthcare. The work of Bickmore et al. (2013) describes a health counselor agent that uses conversations to promote healthy behaviors among its users, specifically performing physical activities, and consuming fruits and vegetables. A study by Fitzpatrick et al. (2019), on the other hand, investigates how Woebot's cognitive behavior therapy skills can help users exhibiting symptoms of anxiety and depression.

As a critic, the agent is assigned to "look over the shoulder" of the users as they perform their tasks, and to offer appropriate advice as the need arises (Terveen, 1995). Critics simulate human problem-solving strategies to identify and inform users of potential concerns during task performance and offer alternative perspectives. The critic must be able to present "a reasoned opinion about a user's product or action", and must "recognize and communicate issues concerning a product" (Fischer et al., 1991a). The critic should also be able to engage the user in problem-solving tasks to resolve issues, and to provide alternative solutions by presenting their advantages and limitations for consideration (Fischer et al., 1991b).

In learning environments, conversational pedagogical agents can vary their level of intelligence and the manner by which they communicate with the learners. These abilities allow them to function not only as experts (Baylor and Kim, 2005), tutors (Graesser et al., 2005) and mentors (Zakharov et al., 2007), but also as learning companions (Cassell et al., 2005) and teachable agents (Zhao et al., 2012). In the latter two cases, the pedagogical agents serve as peers that exhibit collaborative behavior in learning activities through conversations and coaching (Ryokai, Vaucelle and Cassell, 2002), in order to promote a two-way exchange of knowledge.

Because people are inherently social beings, research has also been devoted to developing chatbots that can be interact socially and can engage in empathic conversations. One example is Microsoft Xiaoice, whose main goal is to build a long-term emotional connection with its users. Studies conducted by Zhou et al. (2018) showed that with continued use, users would eventually consider Xiaoice as a friend, frequently engaging in daily conversations with the chatbot by sharing their hobbies and interests.

2.2 Dialogue

A dialogue is defined as a two-way communication activity between two entities, usually with the intent of seeking information, negotiating, persuading, and deliberating in highly contextual tasks and discussions. To facilitate human-agent interaction, both entities must express their intentions through a series of dialogue exchange.

Dialogue exchange is characterized in terms of communicative goals or speech acts. Common speech acts include: inquiry or direct question to solicit additional information; informing to respond to a request for information; elaboration to provide additional definitions or descriptions; justification to explain actions; motivation to persuade someone to carry out an action; exemplification to demonstrate how the task can be carried out, and repair to resolve misunderstanding.

During storytelling, the dialogue exchange between the conversational agent and the user may center around the story itself and its elements, such as characters, setting and plot events, or everyday events that are interesting for the user and that may or may not be directly related to the story. The participating entities can retell story events, describe character attributes, justify character actions and motivations, imagine possible outcomes resulting from character actions or variations of existing stories, and reflect on the relevance of story events to one's daily experiences. This "ability to imagine, solve problems and make decisions", according to Fay (2014), is what makes "story generation play a vital role in intelligence".

3. Conversational Storytelling Agents

Collaborative storytelling entails a two-way exchange of ideas, feedback and suggestions regarding a story. Co-authoring through taking turns to generate story content can also occur during the collaboration. Written text or voice-based interfaces can be used as the medium of communication, both with their respective advantages and constraints as reported in (Ong et al., 2019).

3.1 Orsen, the Storytelling Peer

Orsen is a collaborative storytelling peer with input understanding and text generation abilities designed to help children create their own stories. It uses Google Firebase as the text-based interface, and Google Home and Amazon Alexa as voice-based interfaces. It utilizes a number of dialogue moves to formulate an appropriate response for a given user input. Prompts, such as "*I see, what happens next?*" and "*Tell me more about <character>.*", are used to elicit additional details regarding a story element. Hints, e.g., "*Then Salie went to market.*", allow Orsen to suggest content in order to help a child who is stuck and does not know how to proceed with his/her story.

A collection of commonsense concepts and their relations extracted from children's stories, described in (Ong et al., 2018), is used by Orsen to process user input and to generate a response. This gives children the perception that they are collaborating with an intelligent storytelling peer who understands concepts and their interrelationships normally present in the physical human world. In this way, Orsen can augment the human writer's limited knowledge by offering prompts that could lead to writing ideas.

Because storytelling is a mutual learning process for the entities involved, Orsen also behaves as a teachable agent. It leverages on the principle of learning by teaching (Zhao et al., 2012) when it encounters an unfamiliar concept, i.e., a concept that is not in its knowledge repository. In this situation, Orsen switches to a learner role and generates inquisitive responses such as "*I want to hear more about bandages.*" This knowledge acquisition strategy helps in expanding the agent's knowledge base, but may also lead to learning new assertions that are not necessarily true or acceptable to all users. To address this, Orsen makes a suggestion in subsequent dialogue turns to verify the validity of its acquired knowledge. For example, to verify that a "*swing*" can be found in the "*park*", the agent can ask "*Did she see a swing?*" in response to the child's input text "*The princess walks to the park.*"

Analysis of conversation logs showed that stories shared by children vary depending on their interests and reading habits. Traces of everyday experiences are evident in some stories, expressed through character actions and events. Retelling of popular fables and fairy tales manifest in others. The length of the stories are constrained by the child's linguistic skills and personality. Some children willingly share their stories with Orsen, while others struggle with writer's block. The mode of communication also affected the interaction. Text-based interfaces require correct grammar and spelling, while voice-based interfaces require appropriate pronunciation and accent; otherwise, communication breakdowns would emerge due to Orsen's inability to understand the user's input (Ong et al., 2019).

Most language technologies are based on English, which is also one of the main languages comprising the bilingual nature of most everyday conversations in the Philippines. While Orsen is currently designed to process and generate English text, the presence of common Filipino words, e.g., *tita* (aunt) and *lola* (grandmother), has been detected in a handful of the stories.

Design challenges include establishing the size of the knowledge base to make the agent act intelligently; determining the domain of the assertions comprising this

knowledge base to adequately cover the wide variety of topics that children may share; designating the mode of communication to be used in interacting with children (text or voice); and handling communication failure. Recent advances in language technologies may be able to address speech synthesis issues, produce agent responses free from grammar errors, and correct parse user inputs to properly identify story elements regardless of misspellings and mixed language insertions. A balance must be achieved between a highly intelligent agent capable of following any story topic conceived by the children, and an agent that simply utters “That’s interesting, please tell me more!” (to hide its inability to generate a relevant story content), but still leaves enough space for children to take center stage in the story creation process.

3.2 Towards Emotional Intelligence with Eren

Although stories created with Orsen ultimately remain fictional in nature, children may have drawn from personal experiences to craft their narrative. It was observed that some children shared stories about real-life events that affected them emotionally. On the basis that children can be asked to talk about their emotions freely through storytelling (Denham et al., 1996), this then paves an opportunity for developing emotional intelligence (EI). EI involves recognition of one’s emotions, sensitivity to other people’s emotions, and control of raging emotions.

Built on top of Orsen, Eren is a conversational agent that treats stories shared by children as accounts of their personal experiences. Based on the extracted characters, objects and events from the child’s input, Eren identifies the associated emotion using the OCC Model (Shaikh et al., 2009); validates this with the user; then prompts the user to reflect on his/her actions. Eren adapts Orsen’s dialogue moves using the emotion coaching model of Gottman et al. (1996) to help children understand and resolve their emotions during storytelling. In contrast to Orsen’s free-flowing dialogue controlled by the child, Eren follows a set sequence of phases in its dialogue. This is necessary so that the conversation flows from emotion recognition, to cause identification, to action reflection (“*Do you think what <character> did was right?*”), and lastly to post-evaluation (“*What did you feel after telling me this?*”).

Combining AI with an emotion-aware agent poses certain design challenges. Eren’s emotion recognition ability is not perfect; at times, it may fail to identify the emotion from the child’s input. Thus, it requires a confirmation dialogue (“*You seem <emotion>, is that right?*”) to validate the detected emotion proceeding to the next phase of the conversation. This, however, has led to situations wherein the child indicated that he/she does not feel anything.

Certain pumps from Orsen – particularly those that ask for character and object descriptions, such as “*How old is <character>?*” and “*How big is <object>?*” – were found to distract the child from focusing on the emotional impact of actions and events taking place in his/her story. Because they do not support the agent’s listener role, these types of pumps were subsequently removed from Eren.

Results from end user feedback showed that most children acknowledged Eren as a human (“*I feel like talking to a real human*”), a friend (“*Like a friend who won’t judge you*”) or a therapist (“*therapist vibe*”). Children also identified the preservation of privacy as an important factor in deciding

to share their personal stories with a conversational agent (“*(Eren) won’t share my story with others*” and “*Not a person so it won’t judge me*”). This affirms the findings reported by Fryer and Carpenter (2006) that people tend to be more willing to disclose emotional stories to a conversational agent since it does not lose patience and does not judge.

3.3 Charm, a Chatbot for Joint Story Reading

Soriano and Ong (2016) explored how emotional changes could be elicited from the learner through the use of text-based conversation involving system-generated questions and user responses. By disrupting and shifting the learner’s negative affect to a more positive perception toward learning, the software agent was envisioned to indirectly encourage the learner to complete the learning task at hand. A key design issue revealed in this work, however, was that the virtual peer evokes a conversation only when the learner is taking the reading comprehension test, which is the last part of the learning activity. But end user testing showed that the learners would have preferred an earlier intervention, specifically during the reading task itself. Such results motivated the development of Charm.

Charm is a conversational agent designed to mimic the dialogue exchange that takes place during story reading between an adult and a child. It aims to address three levels of concerns – comprehension, relevance and engagement – which affect the child’s motivation to finish a given reading material. These concerns in turn are supported through three types of dialogues: cognitive dialogue to facilitate comprehension of the text; reflective dialogue to relate story events to the reader’s personal life; and elaborative dialogue to promote engagement with the story by focusing on the key characters and events (Chan and Ong, 2018).

While the user has control over the conversation flow, Charm formulates its response following the mental model and proposition theories of Gunning (1996). Except for *who* and *where* questions, the agent does not give direct answers to other types of questions posed by the user. Instead, Charm rephrases the questions and throws them back to the user, encouraging the latter to seek answers to his/her own question by recalling story elements. Question formulation uses the question circles strategy (Tofade et al., 2013) to construct subject matter questions as well as external reality questions (Chan and Ong, 2018).

Mixed feedback was received from participants (Chan and Ong, 2018). On a positive note, Charm was found to be “*easy to talk to*” despite the need to use the English language which reduces the pleasantness when compared to conversing with a classmate or a friend in mixed languages. The uniqueness of the experience also piqued the interest of the participants, which they said is very different from having a face-to-face conversation. However, the participants’ inherent lack of eagerness to read the story cause major usability issues, in particular, as the agent would keep asking the users to think of the answer on their own; some participants started feeling disappointed and even frustrated with the agent.

As an intelligent agent, Charm’s library is currently limited to only one reading material. It has a manually-built computational model of the major scenes and characters involved in the story. It then uses this model to retrieve answers to *who* and *where* questions, and to formulate

what, why, how and do-you-think questions about character actions and events. Advances in language technology could enable Charm to automatically extract relevant elements from a given input story text.

4. Conclusion and Future Work

Stories abound in everyday human conversations. We explored the potential applications of voice assistants and chatbot interfaces in collaborative storytelling to provide interactive spaces for children to create their own stories, share personal experiences, and introspect/reflect on emotions. We encountered design challenges that affected how children perceived the usability of these agents. Issues concerning effective speech synthesis particularly for bilingual and multilingual users, usage of correct grammar structures in composing written and oral text, and the adequacy and extent of the capabilities of existing language technologies need to be addressed in future work to enrich the child-agent interaction.

5. Acknowledgements

This research is funded through grants from the Department of Science and Technology – Philippine Council for Industry, Energy and Emerging Technology Research and Development (DOST-PCIEERD), and from De La Salle University.

6. Bibliographical References

- Baylor, A.L. and Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *Intl. J. of Artificial Intelligence in Education*, 15(2):95-115.
- Bickmore, T., Schulman, D. and Sidner, C. (2013). Automated interventions for multiple health behaviors using conversational agents. *Patient Education and Counseling*, 92(2):142-148, Elsevier Ireland.
- Blair, D. and Meyer, T. (1997). Tools for an interactive virtual cinema. In Trappl, R. and Petta, P. (eds.) *Creating Personalities for Synthetic Actors: Towards Autonomous Personality Agents*, pages 83-91. New York: Springer.
- Cassell, J., Tartaro, A., Rankin, Y., Oza, V. and Tse, C. (2005). Virtual peers for literacy learning. *Educational Technology special issue on Pedagogical Agents*.
- Chan, L. and Ong, E. (2018). Engaging children in conversations during story reading. In *Proc. of the 26th ICCE Workshop on Innovative Technologies for Enhancing Interactions and Learning Motivation*.
- Coen, M.H. (1995). SodaBot: A software agent construction system, MIT AI Lab, USA.
- Dale, R. (2020). Voice assistance in 2019. *Natural Language Engineering*, 26, pages 129-136. Cambridge University Press.
- Denham, S., Mason, G. and Auerbach, S. (1996). Mother-child dialogue about emotions and preschoolers' emotional competence. *Genetic Psychology Monographs*, 121(3):311-337.
- Fay, M.P. (2014). *Driving story generation with learnable character models*, PhD Dissertation, Massachusetts Institute of Technology.
- Fischer, G., Lemke, A.C., Mastaglio, T. and Morch, A.I. (1991a). The role of critiquing in cooperative problem solving. *ACM Transactions on Information Systems*, 9(3):123-151.
- Fischer, G., Lemke, A.C., McCall, R. and Morch, A.I. (1991b). Making argumentation serve design. *Human-Computer Interaction*, 6(3-4):393-419.
- Fitzpatrick, K.K., Darcy, A. and Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Mental Health*, 4(2).
- Fryer, L. and Carpenter, R. (2006). Bots as Language Learning Tools. *Language Learning and Technology*, 10(3):8-14.
- Gottman, J.M., Katz, L. F. and Hooven, C. (1996). Parental meta-emotion philosophy and the emotional life of families: Theoretical models and preliminary data. *J. of Family Psychology*, 10(3):243-268.
- Graesser, A. C., Olney, A., Haynes, B. C. and Chipman, P. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4):612-618.
- Gunning, 1996 Charm Thomas Gunning. 1996. *Creating Reading Instruction for All Children*, 2nd ed. Allyn & Bacon, Des Moines, IA.
- Ong, D.T., Gilig, L.K., De Jesus, C.R., Albuero, J.B. and Ong, E. (2018). Building a commonsense knowledge base for a collaborative storytelling agent. In Yoshida K., Lee M. (eds.) *Knowledge Management and Acquisition for Intelligent Systems*, LNAI 11016, pages 1-15.
- Ong, E., Albuero, J.B., De Jesus, C.R., Gilig, L.K. and Ong, D.T. (2019). Challenges posed by voice interface to child-agent collaborative storytelling. In *Proc. of the Oriental COCODA 2019*, IEEE.
- Ryokai, K., Vaucelle, C. and Cassell, J. (2002). Literacy learning by storytelling with a virtual peer. In *Proc. of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community*, pages 352-360, Colorado.
- Shaikh, M., Prendinger, H. and Ishizuka, M. (2009). A linguistic interpretation of the OCC emotion model for affect sensing from text. *Affective Information Processing*, pages 45-73, Springer-Verlag London.
- Soriano, Z. and Ong, E. (2016). A conversational agent to shift students' affect state. In M. Baldoni et al. (eds.), *IWEC 2014 / IWEC 2015 / CMNA 2015, Principles and Practice of Multi-Agent Systems*, LNAI 9935, pages 1-12, Springer International Publishing, Switzerland.
- Terveen, L. G. (1995). An overview of human-computer collaboration. *Knowledge-Based Systems*, 8(2-3):67-81.
- Tofade, T., Elsner J. and Haines, S. (2013). Best practice strategies for effective use of questions as a teaching tool. *American J. of Pharmaceutical Education*, 77(7):155.
- Zakharov, K., Mitrovic, A. and Johnston, L. (2007). Pedagogical agents trying on a caring mentor role. In *Proc. of the 2007 Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, pages 59-66. IOS Press.
- Zhao, G., Borjigin, A., and Shen, Z. (2012). Learning-by-teaching: Designing teachable agents with intrinsic motivation. *J of Educational Technology & Society* 15(4):62-74.
- Zhou, L., Gao, J., Li, D., and Shum, H.Y. (2019). The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics*, 1(1):1-62.