Your Turn : Interactive Story Continuing Agent

Report of Surveying Paper Works

Edward Yu-Te Shen

Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
edwards@cmlab.csie.ntu.edu.tw

June 2003

Abstract

Interactive storytelling has drawn researchers' interests in recent years. For educational and entertaining purposes, we wish to build an interactive conversational story continuing agent based on to-date storytelling systems. We introduce our framework and discuss about the possible way to implement the system through introducing related works. Two main features, story understanding and story generation, are focused on, since they from the crucial part of the framework, and we have not seen the integration of these two features into one system yet. Tools to implement the system are also listed for further discuss into the implementation.

Keywords:

Interactive storytelling, embodied conversational agent, story understanding, story generation, narrative intelligence

1. Introduction

Interactive storytelling has attracted more and more research interest in recent years. Researchers believe that interactive storytelling systems will be widely used in the future, since such systems can be applied in educational, entertaining, training and other areas. To-date systems developed diverse in great variety, including character-based storytelling systems [1], interactive story authoring [2], virtual storyteller [3,7,8], virtual peers [3,4], etc. One of the character-based systems, implemented by Cavazza et al. [1], is built under the scenario that users, who is to be a spectator in front of a projection onto a screen of a story play, may interact with the on-screen characters by uttering advice to them or interacting physically with the objects on stage, and therefore influence the plot of the story under development. An interactive authoring interface, namely, TEATRIX, is developed by Paiva et al. [2], and provides a collaborative environment for children to build their own fairy tales and watch the play during and after and creating process.

In Figure 1, Sam [3], introduced by Cassell et al., is an embodied conversational virtual agent, who is a life-sized virtual child projected on the screen behind a toy castle. Cameras and other devices are equipped around so that the system will be able to know whether there is anyone in front of Sam. Sam shares stories stored in the database with children by speaking natural language, invites them to tell stories, and make responses so that it feels like the stories are understood. Stories told by children may begin with less vocabulary, poorer content and shorter length, and become enriched gradually as taking turn telling stories with Sam for several times. Sam's stories influence children's own stories in narrative, structural, and lexical respects, and therefore foster their language skills. Also, "…children get more value from the activity when playmates engage in the activity collaboratively…an embodied conversational character has the educational potential and emotional engagement for this goal…", Sam helps children develop their social and conversational skills as well. However, although reacting as paying fully attention to the child, such as utterance like "Cool! What happen next?", Sam does not really understand what the child is saying, since no natural language understanding is carried out in the system. This leads to failure when children ask Sam explicit questions, or even request Sam to continue the stories they told [5].

Figure 1. Sam, the virtual peer for storytelling introduced by Cassell et al.

As Sam being unable to understand the intentions of the listeners, the children, character-based systems, which provides interfaces for users to interact with the characters or objects in the story after a story begins rather than choosing important features of the story before it, including characters, places, and time in the story, are somewhat constrained by the original story in the system. Interactive story authoring, on the other hand, emphasizes on the collaboration among children rather than children and the virtual conversational agent. The idea of this paper came from the comparison among these previous works. In the following sections, by introducing other related works, we will discuss about the possible way of building a system that is able to make up stories collaboratively with the users. More specifically, our goal is an embodied conversational agent who can understand stories told by the users, continue to expand them, and also answer questions about them.

2. System Overview

Cassell et al. proposed an embodied conversational agent, REA [6], named after the real estate agent, with the architecture adequate for the engagement of daily conversation. Figure 2 shows the architecture of REA. The input manager of REA is responsible for integrating input data from devices of various modalities, including speech, gestures, head nods and eye gaze, into a single semantic representation. The core module, dividing in three sub-modules including understanding, decision, and generation module, plays the central role in the conversation process. In this module, proper reaction is produced accordingly after propositional and interactional information, which are information about the content of the conversation and the discourse itself, respectively, are understood. Finally, in the output stage, the action scheduler schedules and synchronizes acts of all the modalities to be performed, and pass them to the output devices such as speech synthesizer and animation.

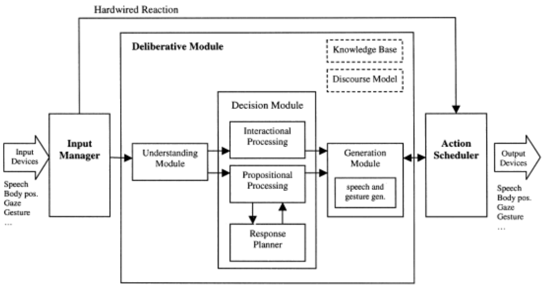

Figure 2. The software architecture of Rea

The architecture above provides a good example for the implementation our framework, while some functions need to be revised, since the conversation in REA's case is limited in a small domain, that is, information of real estate. Firstly, the understanding module. While the speech recognition and the natural language understanding of REA may be fine enough to recognize words and sentences in a story, they are incapable to understand the plot, which depends on the interactions, emotions, and personalities of characters, temporal information, and causal information. Narrative effect is not to be dealt with properly, since the relationships among sentences are too complex comparing with the simple speech acts in conversation with REA. The lack of common sense of the system also leads to failure of making inferences. (For example, after eating a great meal, it is of high possibility that one may not be hungry; A man going to the airport with a baggage may be going to somewhere by plane.) The incapability of fully understanding stories of the understanding module reasonably leads to failure making decision in the decision stage. But what if the understanding module is well established? The decision module of REA is implemented as a rule-based expert system, which means REA response to users' acts under predefined rules. In the storytelling case, since stories vary so much with each other, defining the rules for all the situations our agent may face will be a heavy and even everlasting work. Also, there might be more than one goal under a single circumstance, which increases the difficulty implementing storytelling as a single rule-based expert system.

To overcome these problems, two important features need to be embedded, namely, story understanding and story generation. The goal of story understanding is to interpret input stories of speech, text, drama script or other forms into relationships of the characters with temporal and causal information. As for story generation, it aims to create story in natural language, carrying narrative, dramatic and other effects. Both of these features are necessary for reconstructing the architecture of the agent we proposed. We now introduce more about the previous works in these two areas, and discuss them with some other related works.

3. Understanding The Narrator's Story

3.1 Plan Recognition

With plan recognition, researchers try to build agents who can understand people or other agents' speech or acts. Applications include user modelling like intelligent help systems or tutoring agents, multi-agent interaction, natural language applications like story understanding, machine translation, dialogue systems, etc [9]. Plan recognition is typically divided into two types, *keyhole recognition* and *intended recognition. Keyhole recognition* refers to recognition where the acting agent or user is unaware of the happening of the recognition, while under the *intended recognition* case the acting agent or person makes efforts to clarify his/her plan to the recognizing agent.

Early works on plan recognition in discourse based on AI planning systems since 1980. In 1990, Lochbaum et al. proposed a model of collaboration called SharedPlan, and presented an algorithm based on this model a year later, which is of intended recognition [10]. Another approach developed by Carberry [11], uses Dempster-Shafer (D-S) belief functions. In these works, the idea of uncertainty had not been commonly adapted in discourse recognition, until the approach using Bayesian network proposed by Charniak et al. [12]. Charniak argued that the problem of plan recognition, "is largely a problem of inference under conditions of uncertainty" and indicated problems of plan recognition systems, particularly on discourse recognition, without considering uncertainty. From then on, Bayesian network has been commonly adapted while discourse recognition problems are considered.

3.2 Charniak's Work : Using Bayesian Model

Charniak et al. described a set of rules translating plan recognition problems into Bayesian networks, and implemented a natural language story understanding system, WIMP3, to prove that their approach outperformed the other systems. With a knowledge-base, WIMP3 dynamically constructs its network, which shows the plan recognition of the plot of the story. The network shown in Figure 3 is constructed by WIMP3 after seeing "Jack went to the liquor-store". If the system produces more than one hypothesis during the recognition, the hypotheses will compete with each other. Figure 4 shows the competition between shopping and robbery after "He pointed a gun at the owner", was added. Either each of the states or the outcomes of the competitions in WIMP system depend on possibilities of the candidates. For the example in Figure 3, it is also possible that Jack goes to the liquor-store for robbery, but in the network we do not see that it is considered as a possible plan, since the possibility of someone who goes to a liquor-store for robbery is too small. After the latter sentence is revealed to the system, the robbery plan is added into the network as a candidate, because the possibility of robbery is much higher in the case that someone points his gun at the owner.

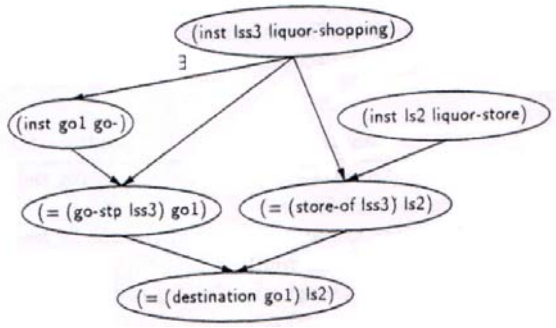

Figure 3. The Second Bayesian network of reconized plan

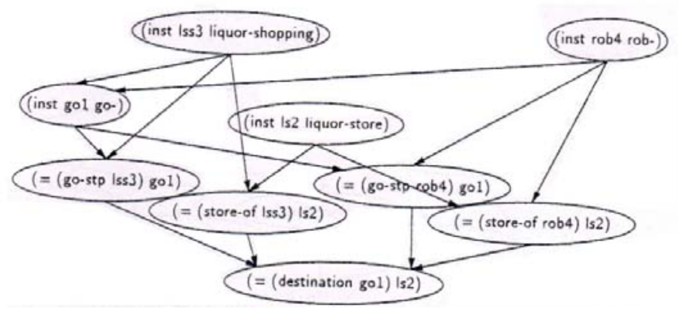

Figure 4. The Second Bayesian network of reconized plan

The mechanism based on Bayesian networks easily prevents the recognition process from the problem faced by rule-based system when multiple hypotheses are produced. In the work of Charniak et al., it is proved to be a more proper framework for story understanding than other heuristics. Not only are the authors themselves confident with the Bayesian model approach, as "…It seems unlikely to us that an approach which did not use an uncertainty calculus would have much chance to succeed…", Blaylock, when arguing that most plan recognition systems only work on a complete speech act rather than beginning the recognition as long as the utterance proceeds, which may lead to serious problems in real-time interaction, suggested that "…There is hope of a possible solution, which may be able to build on the belief network plan recognition system of Goldman and Chariak…their system seems to be already support a limited from of incremental plan recognition…" [9]. The importance of Bayesian network in story continuing system is thus emphasised on, since continuous recognition has to be performed during the conversational storytelling processes.

3.3 Reading Comprehension Systems

Reading comprehension systems are automated reading systems that can receive arbitrary text input and answer "WH" questions, that is, Who, What, When, Where, and Why, about it, just like what students do in a reading comprehension test. We introduce two of the modern reading comprehension system here, since when the story continuing agent is asked these questions about the story told by the narrator, for example, a child in front of the system, it can be analogous with that the system is taking a comprehension test. DeepRead [13] and Quarc [14] both make use of computational linguistic techniques and WordNet [15] system to extract information from text input. Although they are both far from practical use as a question answering module embedded in our framework (with overall performance no more than 60%), there might be some advantages combining them with the projects in the previous sections.

4. Planning & Story Generation

4.1 Structure-Based Storytelling systems

TALE-SPIN [16], a story generating program early in 1977, provide a schema of computational storytelling, and influenced later storytelling systems including MINSTREL [17], HOMER [18], and that described by Clark [19], etc. Here is a sample story created by TALE-SPIN in [17]:

> *John Bear is somewhat hungry. John Bear wants to get some berries. John Bear wants to get near the blueberries. John Bear walks from a cave entrance to the bush by going through a pass through a valley through a meadow. John Bear takes the blueberries. John Bear eats the blueberries. The blueberries are gone. John Bear is not hungry.*

Although stories created by TALE-SPIN do not read like *stories*, that is, they lack many properties a story should possess, TALE-SPIN forms story in a reasonable way. TALE-SPIN is somewhat interactive, since it asks users to establish the characters in the stories, and let the users to choose about what happen at important points of the stories, while which is also the weakness of TALE-SPIN [18].

Stories generated by MINSTREL satisfy Thematic, Drama, Consistency, and Presentation goals. In other words, the stories are plausible, believable, of high artistic quality, and well presented to the readers. A moral like: "Done in haste is done forever" would even be concluded from the story. The core module of MINSTREL performs the storytelling process by selecting a goal from a pool of goals, executing it, and iterate until there are no goals in the pool, which is the end of the story. This can be divided into two process, featured by Figure 5.

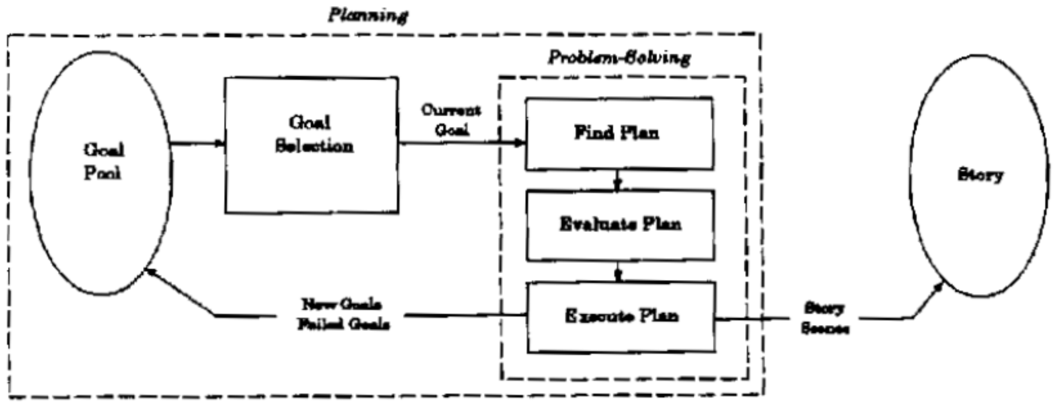

Figure 5. Planning module of MINSTREL

Clark [19] proposed a framework based on TAIL-SPIN, both of them are implemented as STRIPS-like planners. However, in his generalized work, the planner can explore possible branches, and sophisticated approach allows conditional plans. Meanwhile, the augmentation of multiple agents and goal revision also reduces undesirable outcomes.

Storytelling systems can be roughly divided into two categories: structure-based and environment-based [18]. TAIL-SPIN, MINSTREL, and Clark's work all belong to the former one and follow storygrammer or have strong structures. For the story continuing agent, these approaches may be fine to be integrated as a module, since the data it receives, which comes from the story understanding module, should be translated into relationships among the characters and objects in the stories, or, further more, a story structure which needs to be expanded. If the story generation heuristics above could be designed to regard the input data from the understanding module as incomplete goals within itself, the core module of the story continuing agent will be fulfilled.

4.2 HOMER

HOMER and is nearly the most recent story generating system, proposed by Konstantinou [18]. What makes HOMER different is that (1) stories generated by HOMER possesses *point-of-view* and *style* properties that can hardly seen in almost any of the previous works, and (2) one of HOMER's objective it is to be integrated with CONFUCIUS, a multimedia story presentation system [8], so that the outcome may be a multi-modal presentation from simply text input. This will be somewhat close to our framework, while several differences exist, including: (1) HOMER cannot receive speech input, (2) HOMER cannot interact with the users, (3) HOMER is not designed to be able to understand stories. Konstantinou's paper describing HOMER provides a clear comparison among storytelling systems, which is of great help clarifying differences among all these systems.

4.3 Reader-Based Story Generation

Bailey introduced an idea [20,21] different from all the other researchers, which is regarded as helpful to our framework. He argued that "A story needs a reader", and suggested that the system should consists of an "author model" and a "reader model" to perform a "virtual telling" of a story. Although in our framework, there already exists one or more listeners when the stories are being told, we believe that Bailey's model will lead to stories that interest the listeners much better, since the generating process is extremely related to the questions and expectations hold by the reader model in the system. Figure 6 shows the architecture of Bailey's system.

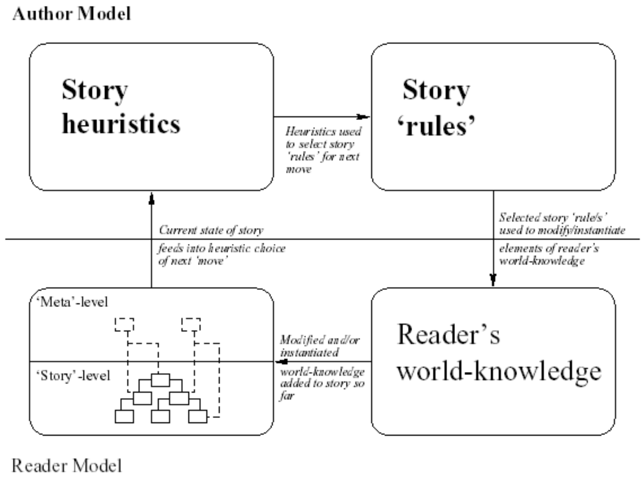

Figure 6. The architecture of Bailey's approach

5. Tools

Many tools are needed to implement our framework, and are listed here.

a) **Speech recognition:** SUMMIT or IBM Via Voice

b) **Speech Synthesiser:** Microsoft Whisper or Festival

c) **Natural Language Generation:** SPUD

d) **Image Processing:** STIVE

e) **Common Sense Database:** ThoughtTreasure or Open Mind Common Sense

6. Conclusion

We introduce interactive storytelling systems and discuss about a framework that may be extended from these previous works, which is to build a interactive conversational story continuing agent, who is able to understand story told by users, expand them into more complete ones, and tell them multi-modally. The most important features, the story understanding module and the story generating module, are discussed through reviewing previous works. Tools which can be used in the implementation of the system are also introduced, and the future work may be more specific details of the implementations.

References

1. Cavazza, M., Charles, F. and Mead, S.J., 2002. "Interacting with Virtual Characters in Interactive Storytelling". *Proc.of ACM Joint Conference on Autonomous Agents and Multi-Agent Systems*, pp. 318-325.
2. A. Paiva, I. Machado, and R. Prada, 2001. "Heroes, Villains, Magicians, …: Dramatis Personae in a Virtual Story Creation Environment". *Proc. of IUI 2001*. ACM, 2001.
3. J. Cassell, M. Ananny, A. Basu, T. Bickmore, P. Chong, D. Mellis, K. Ryokai, J. Smith, H. Vilhjálmsson, and H. Yan, 2000. "Shared Reality: Physical Collaboration with a Virtual Peer", *Proc. of CHI 2000*. ACM, 2000.
4. J. Cassell and K. Ryokai, (to appear)"Making Space for Voice : Technologies to Support Childern's Fantasy and Storytelling", Personal Technologies.
5. Justine Cassell (under review), "Towards a Model of Technology and Literacy Development: Story Telling Systems".Also published as Tech Report ML_GNL_01-1.
6. Justine Cassell, 2000. "More Than Just Another Pretty Face: Embodied Conversational Interface Agents", *Communications of the ACM*, Vol. 43, No. 4, pp. 70-78, 2000.
7. M. U. Bers and J. Cassell, 1998. "Interactive Storytelling Systems for Children: Using Technology to Explore Language and Identity", *Journal of Interactive Learning Research*, Vol. 9, No 2, pp. 183-215, 1998.
8. Minhua Eunice Ma, 2002. "CONFUCIUS: An Intelligent Multimedia Storytelling Interpretation and Presentation System", First Year Report, School of Computing and Intelligent Systems, University of Ulster, Magee.
9. Nate Blaylock, 2001. "Retroactive Recognition of Interleaved Plans for Natural Language Dialogue", Technical Report 761, The University of Rochester, Computer Science Department, Rochester, New Work.
10. Karen E. Lochbaum, 1991. "An algorithm for plan recognition in collaborative discourse", *Proc. of the 29th ACL,* pp. 33-38, 1991.
11. Sandra Carberry, 1990. "Incorporating default inferences into plan recognition", *Proc. AAAI-90,* Boston, MA, 1990.
12. E. Charniak and R. P. Goldman, 1993. "A Bayesian model of plan recognition", *Artificial Intelligence*, Vol. 64, pp. 53-79, 1993.
13. L. Hirschman, M. Light, E. Breck, and J. Burger, 1999. "Deep Read: A Reading Comprehension System", *Proceeding of the 37th Annual Meeting of the Association for Computational Linguistics*, 1999.

14. E. Riloff and M. Thelen, 2000. “A Rule-Based Question Answering System for Reading Comprehension Tests”, *ANLP/NAACL-2000 Workshop on Reading Comprehension Tests as Evaluation for Computer-Based Language Understanding Systems,* 2000.
15. G. A. MILLER, 1995. “WordNet: A lexical database for English”, *Communications of the ACM*, Vol. 38, No. 11, pp. 39–41, 1995.
16. J. R. Meehan, 1977. “TALESPIN, an interactive program that write stories”, *IJCAI-77*, pp. 91-98.
17. S. R. Turner, 1992. “MINSTREL: A computer model of storytelling and creativity”, Computer Science Department Technical Report, University of California, Los Angeles, CA.
18. D. N. Konstantinou, 2001. “HOMER: An Intelligent MultiModal Generation System”, Research Plan, Faculty of Informatics, University of Ulster, Magee, Londonderry.
19. Peter Clark, 1999. “Story Generation and Aviation Incident Representation”, Working Note 14, Knowledge Systems, Applied Research and Technology, The Boeing Company.
20. Paul Bailey, 1999. “A Reader-Based Model of Story Generation; or `Stories: they're not what you expected'”, AISB`*99 Symposium on Creative Language: Humour and Stories,* University of Edinburgh, April 1999
21. Paul Bailey, 1999. “Searching for Storiness: Story-Generation from a Reader's Perspective”**,** *AAAI Fall Synposium on Narrative Intelligence,* Nov. 1999.
22. M. Maragoudakis, A. Thanopoulos and N. Fakotakis, “Statistical Decision Making applied to Text and Dialogue Corpora”, *Text, Speech and Dialogue, Proc. of 5th International Conference, TSD 2002,* Brno, Czech Republic, September 9-12, 2002
23. N. Lesh, C. Rich, C. L. Sinder, 1999.“Using Plan Recognition in Human-Computer Collaboration” *Proc. of the Seventh International Conference on User Modeling,* pp. 23-32. New York.
24. Karen E. Lochbaum, 1998. “A Collaborative Planning Model of Intentional Structure”, *Computational Linguistics*, Vol. 24, No. 4, Dec. 1998.
25. Lyn Pemberton “A Modular Approach to Story Generation”, *Proc. of the 4th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 217-224, 1989.
26. I. Machado, A. Paiva, P. Brna, “Real Characters in Virtual Stories”, *Virtual Storytelling, Proc. of International Conference ICVS 2001 Avignon*, France, September 27-28, 2001,
27. Clark Elliott, 1998. “*Story-morphing* in the Affective Reasoning paradigm: Generating stories semi-automatically for use with ‘emotionally intelligent’ multimedia agents”, *Proc. of the Second International Conference on Autonomous Agents*, New York: ACM Press.
28. B. Hayes-Roth, L. Brownston, E. Sincoff, “Directed Improvisation by Computer Characters”, Technical Report: KSL-95-04, Knowledge Systems Laboratory, Stanford University.
29. Open Mind Common Sense: http://openmind.media.mit.edu/
30. ThoughtTreasure: http://www.signiform.com/tt/htm/tt.htm
31. Story Understanding Resources: http://xenia.media.mit.edu/~mueller/storyund/storyres.html