MAKING THE CASE FOR NLP IN DIALOGUE SYSTEMS FOR SERIOUS GAMES

Brad Hennigan

Institute for Interactive Arts and Engineering (IIAE), University of Texas at Dallas
bhennigan@utdallas.edu

Abstract. As computational capability continues to increase, the tools available to designers of digital games have become more robust, allowing high fidelity graphics and sound to become common, and resulting in a market saturated with kinetic-based games. However, consumers and educators are eschewing such games for more complex and immersive stories, the creation of which has proven a difficult mountain for designers to climb. A central reason is that story-immersive games rely on dialogue between the player character (PC) and non-player characters (NPCs), the writing and coding of which is time consuming and inefficient. This paper documents the author's experiences with complex, branching dialogue systems, and examines the possibility of system automation through natural language processing (NLP).

Keywords. serious games, NLP, NLU, NLG, dialogue systems

1 Introduction

It has been well established that narrative in games and simulations is essential to player immersion and positive learning outcomes.[1-4] Yet inconsistencies within narrative and dialogue are all too common. Game developers rarely release games with major, known graphics issues, but narrative problems are often overlooked and accepted. This acceptance is at odds with the goals of serious and educational games, for "if we are to subscribe to the notion of using serious games to better immerse, transport, motivate, and educate the player within a certain educational domain, then we must also consider the potential impact of including a story with incomplete or incongruent information within a specific game environment".[5]

Many training and educational serious games center around social interaction, and by nature require human verbal and nonverbal communication replication as a game mechanic. This paper examines the complexity of creating branching narrative and dialogue systems for serious games, and the potential for heightening player agency and immersion by automating these processes with natural language processing.

2 The Game Writer's Conundrum

Writing for games can be tricky, particularly if the game world is "open," meaning the player is free to roam anywhere within the boundaries of the world and speak to whomever she chooses. Narrative game designer Mary DeMarle describes the potential for the writer's conundrum in such a world:

> A writer walks into a design meeting at We Make Fantasy Games, Inc., an independent game developer….She starts her pitch. She's barely five minutes into her description of how the hero stumbles upon a nondescript cabin in the woods and goes inside…, when the lead game designer interrupts her.
>
> "Hold on. What if the player doesn't go in the cabin?"
>
> "He has to go in the cabin."
>
> "Well, what if he doesn't? What if he decides to check out the next town first? Or what if he decides to blow the cabin up?"
>
> "…Why would he do that?"
>
> "Because he can."[6]

Clearly DeMarle's upstart writer had not thought about the myriad of possibilities available to the player at any given moment in an open-world game, nor had she begun to imagine how to accommodate those possibilities.

2.1 The Role of Memory

Human communication can be viewed as intrapersonal, interpersonal, and public/socio-cultural interdependent systems.[7] Memory is integral to these communication systems; they require the memory components of retention and recall to properly function. Within these systems, memory serves three broad purposes:

1. It acts as a repository for the experiences, concepts, and words (the rhetorician's "available means of persuasion") which are the raw materials of speech inventions.
2. It acts, in connection with processes of thinking and reasoning, as a setting for linking experiences and concepts with words to produce oral expression, or the generation and transmission of a message.
3. Perhaps most importantly, it acts as a vehicle for interpreting and evaluating messages, and for determining how one should respond to them.[8]

One of the most successful commercial games of 2011, *Skyrim*, chose to depart from purely kinetic game play with a well-constructed, character-driven branching narrative. However, the overlooking of narrative issues is prevalent in the game, particularly in regards to dialogue and character memory. For example, ten minutes into the game experience, the player encounters a master of torture deep in the dungeons of Helgen's Keep. Upon speaking to him, the player is told to leave the torturer alone. It is clear the conversation is over, and the player moves on.

If the player then speaks to the torture master again, not only does the torture master give no indication of remembering the player, but in fact repeats the exact same dialogue, with the exact same intonation, gestures, and facial expressions as the first encounter. Dynamic immersion is lost; as Wizards of the Coast Senior Developer Frank T. Gilson states, "bad dialogue will jar the participant out of the experience. Great dialogue will engage and motivate. The challenge is to *prepare* dialogue to account for the interactivity of gameplay".[9]

2.3 Branching Dialogue Systems

Ruth Aylett and her colleagues have established that "if characters are to interact intelligently and meaningfully among themselves, their different potential relationships with each other must be thought through and their place in the world must be clearly established".[10] To achieve this level of meaningful interaction, researchers at the University of Texas at Dallas Institute for Interactive Arts and Engineering (IIAE) utilize a branching complexity within gaming narratives and conversations, termed *dynamic immersion*, which plans for all potential player movement and conversations throughout the game world, and simulates the feel of self-determination.

IIAE conducted extensive research into a viable methodology for mapping, writing, and coding dynamically immersive, complex, non-linear branching dialogue. Preparation is the key: a game narrative utilizing ten characters and a three-choice branching system can have as many as three to the tenth power, or 59,049, unique conversation branches. While this total will not likely be reached due to dead-end or win-state paths, the writing of half this number could easily be required.

Implementation of IIAE's methodology sparked a keen interest in the possibility of automating some of the processes involved. Several research outcomes support the need for such automation:

- *Volume of conversations branches*: The number of unique conversation branches for a scenario with eight Key NPCs was estimated at $3^8 \div 2$, or $6561 \div 2 = 3280$. The actual number of unique conversation branches realized was 2,263. One of the Key NPCs was a "help agent" that had direct conversations with the player, which accounts in part for the reduction in unique conversation branches.
- *Resource requirements of methodology*: The methodology required 400 man hours to map, write, and code 45 minutes worth of game play, which is highly inefficient. In comparison, *Skyrim's* main quest takes approximately thirty hours of game play to complete.[11] If the game were written and coded using the IIAE

methodology, it would have taken 16,000 man hours, or a team of three writer/coders almost two and a half years to complete.

3 NLP in Dialogue Systems: Input vs. Output

Determining the feasibility of using NLP in dialogue systems for games begins with separating NLP into two of its functions: natural language understanding (NLU), which aims to simulate human understanding of language[12]; and natural language generation (NLG), which aims to compute natural language sentences or texts that convey a given piece of information to an audience.[13] In other words, the understanding by the system of the PC's utterances, or input (NLU); and the generation by the system of an NPC's responses, or output (NLG).

3.1 NLU and PC Input

Traditionally, dialogue systems in games require manually writing both the PC and NPC dialogue. However, the use of NLP, and specifically NLU, in gaming is not without precedent. Researchers Michael Mateas and Andrew Stern created Façade, an interactive drama that integrates real-time, autonomous agents, drama management, and broad, shallow support for natural language understanding and discourse management.[14] Façade allows the PC to type an utterance, thus eliminating the need to author PC dialogue. The PC input is then mapped into one of a pre-defined set of discourse acts representing the pragmatic effects of the utterance, such as praise, criticism, flirtation, and provocation. This determination then triggers an NPC response, drawn from "stacks" of pre-defined context-appropriate responses.[14-15]

While a significant achievement in incorporating NLU into gaming, the pre-definition required in the Façade NLP system took approximately 3 person years to complete, and as such does not represent an efficient means of automating PC dialogue. The designers also realized an expected 30% failure rate in NLU.[14-15] However, current advancements in NLU research could combine with aspects of the Façade architecture to produce a more efficient system usable in serious games.

3.2 IBM's Watson

On February 16, 2011, after three grueling days of answers and questions, *Jeopardy!* champion Ken Jennings wrote on his digital answer pad, in addition to his correct Final Jeopardy question, "I, for one, welcome our new computer overlords." He was referencing the winner of this particular tournament: IBM's supercomputer Watson, which had handed Mr. Jennings and another top *Jeopardy!* winner, Brad Rutter, a sound defeat.[16]

The team of 20 IBM scientists behind Watson had a formidable task: to advance and incorporate QA technologies such as parsing, question classification, question decomposition, automatic source acquisition and evaluation, entity and relation detection, logical form generation, and knowledge representation and reasoning into a single entity; requiring confidence and precision in answering complex natural language questions over a broad domain of topics, all within an average answer time of three seconds per question. They dubbed the project DeepQA, a "massively parallel probabilistic evidence-based architecture...[using] more than 100 different techniques for analyzing natural language, identifying sources, finding and generating hypotheses, finding and scoring evidence, and merging and ranking hypotheses."[17]

Watson's near-instantaneous generation of semantically and contextually accurate questions to Alex Trebek's answers is a great advancement in the NLU aspect of NLP; as long time NLP researcher Ray Kurzweil observed, "it certainly demonstrates the rapid progress being made on human language understanding."[18] Using a system such as Watson in every-day gaming applications, however, is problematic, with the most glaring issue being its hardware requirements. Running on 2,880 CPUs in a massively parallel system, all of Watson's lookup of reference content and analytical operations use structures optimized for in-memory manipulation, requiring 16 terabytes of RAM.[19] As such, Watson is also prohibitively expensive.

A reason behind its massive design is the need to access and process the multiple knowledge bases required of a general knowledge trivia game such as *Jeopardy!* By contrast, the typical serious educational game attempts to train the user on a comparatively narrow band of knowledge. A logical relationship seems to exist between the system's size and the number of knowledge bases needed to properly execute the desired outcome. A serious game that teaches, for instance, proper use of a pediatric respirator, and as such requires access to significantly fewer knowledge bases, should by extension require far fewer processors and RAM to achieve its desired outcome.

It is possible that Watson's architecture, running on a pared down system designed to process within a narrow band of knowledge bases, could combine with predefined NPC responses in "stacks" such as those in the Façade system; resulting in a hybrid system that relies solely on deep QA capability to process PC input, while relying on context-appropriate authored responses to generate NPC output. Such a system would eliminate a significant amount of pre-definition in traditionally-authored games, and in systems such as Façade would likely result in a much lower NLP failure rate.

3.3 NLG and NPC Responses

While Watson did generate contextually accurate natural language questions in response to the *Jeopardy!* answers, it was only required to do so within a limited framework: all questions and answers in the game are one sentence long, with the answer in the form of “who/what is/are X?”[19] In the serious games construct described above, such relative simplicity in an output framework may be acceptable, but it is likely that more complexity in the generation of NPC responses would be required.

A surge in research in the past decade has centered on NLG systems that utilize automated planning, which seeks general techniques for efficiently solving the action sequencing NLG requires.[13] Researchers Alexander Koller and Ronald Petrick recently put the latest generation planners to the test specifically in the task of NLG. They found mixed results, with the best planners able to easily handle the search requirements of NLG, while spending 90% of runtime on preprocessing to analyze the problem domain in support of the search. They concluded that the planners tested were generally too slow to be used in real NLG applications.[13] It is possible that future increases in computational capability will make real-time generation of complex NPC dialogue possible, but current limitations relegate such generation to a simple framework such as that found in Watson’s *Jeopardy!* game.

4 Conclusion

The impact of semantically and contextually accurate natural language processing is clear: no longer will the PC be restricted in dialogue choices to what the game’s writers have deemed important to moving the game’s plot forward, but rather she will be able to simply talk to the game and its characters, and they will respond appropriately. This will open up the narrative in games to the same potential level of high fidelity that their graphical representations have enjoyed for decades. Such a system would greatly reduce the need for writer/programmers to manually generate PC dialogue, and would likely infuse the gaming experience with a much greater degree of dynamic immersion, as the player would truly feel she is creating her own story as the game unfolds.

References

1. Fisch, S.M.: Making Education Computer Games Educational. In: Proceedings of the 2005 Conference on Interactive Design and Children. ACM, New York (2005)
2. Marchiori, E. J., et al.: A Visual Language for the Creation of Narrative Educational Games. J. Vis. Lang. & Comp. 22.6, 443-52 (2011)
3. McQuiggan, S.W., et al.: Story-based Learning: The Impact of Narrative on Learning Experiences and Outcomes. Lec. Notes Comp. Sci. 50.91, 530-39 (2008)
4. Waraich, A.: Using Narrative as a Motivating Device to Teach Binary Arithmetic and Logic Gates. In: Proceedings of Ninth Annual SIGCSE Con-

ference on Innovation and Technology in Computer Science Education. ACM, New York (2008)
5. Cannon-Bowers, J., Bowers, C. eds.: Serious Game Design and Development: Technologies for Training and Learning. Information Science Reference, Hershey, PA (2010)
6. DeMarle, M.: Nonlinear Game Narrative. In: Bateman, C. (ed.) Game Writing: Narrative Skills for Videogames. Charles River Media, Boston (2007)
7. Mortensen, C. D.: Communication: The Study of Human Interaction. McGraw-Hill, New York (1972)
8. King, C. P.: A Functional Model of Memory in Communication. In: Annual Meeting of the International Communication Association. Education Resources Information Center, New Orleans (1974)
9. Krawczyk, M., Novak, J.: Game Development Essentials: Game Story & Character Development. Thomson Delmar Learning, New York (2006)
10. Aylett, R., et al.: Unscripted Narrative for Affectively Driven Characters. Comp. Graph. Appl.. 26.3, 42-52 (2006)
11. Hill, O.: E3 2011: Skyrim's Main Quest 30 Hours Long. Additional Content 'Two to Three Hundred' More. PC Gamer. 8 June (2011)
12. Sabah, G.: Natural Language Understanding, Where Are We Going? Where Could We Go? The Computer Journal. 54.9, 1505-1513 (2011)
13. Koller, A., Petrick, R.: Experiences with Planning for Natural Language Generation. Computational Intelligence. 27.1, 23-40 (2011)
14. Mateas, M., Stern, A.: Natural Language Understanding in Façade: Surface-text Processing. In: Proceedings of the Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE). Darmstadt, Germany (2004)
15. Mateas, M., Stern, A.: Structuring Content in the Façade Interactive Drama Architecture. In: Proceedings of the First Artificial Intelligence and Interactive Digital Entertainment Conference (2005)
16. Markoff, J.: Computer Wins on 'Jeopardy!': Trivial, It's Not. The New York Times, 16 Feb. (2011)
17. Ferrucci, D., et al.: Building Watson: An Overview of the DeepQA Project. AI Magazine. 31.3, 59-79 (2010)
18. Kurzweil, R.: The Significance of Watson. Kurzweil Accelerating Intelligence. Kurzweilai.net, 13 Feb. (2011)
19. Nadkarni, P., Ohno-Machado, L., Chapman, W.: Natural Language Processing: An Introduction. J. Am. Med. Inform. Assoc. 18, 544-51 (2011)