Say Anything: A Massively Collaborative Open Domain Story Writing Companion

Reid Swanson and Andrew S. Gordon

The Institute for Creative Technologies, The University of Southern California
Marina Del Rey, Ca, 90292 USA
{swansonr,gordon}@ict.usc.edu

Abstract. Interactive storytelling is an interesting cross-disciplinary area that has importance in research as well as entertainment. In this paper we explore a new area of interactive storytelling that blurs the line between traditional interactive fiction and collaborative writing. We present a system where the user and computer take turns in writing sentences of a fictional narrative. Sentences contributed by the computer are selected from a collection of millions of stories extracted from Internet weblogs. By leveraging the large amounts of personal narrative content available on the web, we show that even with a simple approach our system can produce compelling stories with our users.

Keywords: Interactive Storytelling, Interactive Fiction, Collaborative Writing, Social Media, Information Retrieval.

1 Introduction

Interactive storytelling and interactive fiction on the computer have blossomed into active fields of research. Early work in these fields provided access to virtual worlds through textual descriptions, for example *Adventure* [1]. Although the level of sophistication varied, natural language was used to interact and learn about the underlying narrative structure. Today most state-of-the-art systems provide fully immersive 3D environments to depict what was once only abstractly available through textual responses [2][3][4][5]. Although the interface of modern systems is significantly different than their predecessors, many of the underlying goals are the same. Ultimately, one would like a system in which two competing and somewhat antithetical propositions are upheld. The human user should be free to explore and do anything in the world, receiving appropriate responses as they go. At the same time, a coherent underlying narrative should also be maintained in order to provide structure and meaning [6].

While much of the interactive media world has embraced advanced graphics technology for visualizing stunning virtual worlds, there are many reasons why purely textual works are still a valuable form of entertainment and a testbed for research. Although often scorned by both game developers and literary experts, Montfort [7] argues for the importance of textural interactive fiction on several levels. These include being a platform for computational linguistics research and improving language skills and reading comprehension for the user. There also seems to be a void in the entertainment landscape between language games such as crossword puzzles and scrabble and graphic-rich video games.

Collaborative fiction is a particularly interesting literary exercise and topic to pursue because it is one of the few exceptions in traditional media in which the reader can also be an active participant in the shaping and unfolding of the narrative. Unlike traditional literature, which is perceived as a very personal act of creation with strong (single) authorship, collaborative works share the role of author amongst many individuals. Each individual contributes a portion of the story but the resulting piece has a life beyond any one of the creators. Despite the prominence of single-author assignments, the history of collaborative fiction is as long as ancient texts, such as the *Illiad* and the *Old Testament*. Collaborative fiction also has avenues for generating narrative works. For example, role playing games such as Dungeons & Dragons [8] are often seen as a process to generate narratives that arise through each character's decisions.

Computers and the Web have greatly facilitated communication amongst large numbers of people, which has led to an explosion of collaborative fiction between individuals. Not only has the Internet greatly assisted and expanded upon older collaborative writing genres, it has also spawned a variety of unique genres in its own right. Wikinovels [9] and hypertext [10] are a few of the more popular examples. Rettberg [11] gives a more detailed introduction to these and other collaborative fiction genres.

The type of collaborative fiction we are interested in throughout this paper, however, is much closer to a traditional notion where people take turns writing segments of a story. This process can be an open-ended endeavor with no restrictions on content, length or structure but in many cases rules can be specified to constrain the work. Rules can pertain to the syntax of the story (such as the maximum number of words per turn or fixing the point of view of the story). Rules can also enforce semantic constraints like specifying a particular genre or requiring characters to have certain characteristics. Other types of rules can also be specified and are only limited by the imagination of the designers.

An important theoretic belief of this work is that when individual contributions are put together, a collective wisdom emerges that produces interesting relationships and properties that can transcend single authorship. With the introduction of the Web, vast amounts of people have easy access to each other with inexpensive means of publication. This has pushed the ideology of collective wisdom farther than ever before.

In this paper we will present a new type of interactive storytelling system that blurs the line between traditional interactive fiction and collaborative writing. We do this by leveraging the enormous amounts of content authored by ordinary people with a system that uses this content to take turns writing sentences with a single user. Potentially, millions of bloggers work together with the user of our system in a collaborative writing process to construct a new narrative work. We call our system *Say Anything* because, in the spirit of Bates' vision "Go anywhere and do anything" [2], the user is completely unconstrained in the topics about which they can write.

2 Say Anything

The primary virtue of our system is the simplicity in which our solution achieves its results. When the user contributes a sentence to the emerging story, we simply identify the closest matching sentence in a large database of stories and return the following sentence of *that* story. With virtually no other preprocessing or modeling, relatively high-quality stories can emerge. For example, consider the following story segment created by a user of the *Say Anything* system:

> *You'll never believe what happened last night! Leigh laughed at my joke but I couldn't help but think 'liz would have laughed harder.' The joke wasn't very funny in a "ha ha" kind of way.. It wasn't anything like that, I thought he was going to give me a good night kiss but he ended up licking my cheeck," she declared. It made me sneeze and snort out loud. And now my nose hurts from the snorting.*

Instead of relying heavily on narratology and other narrative theories often used for interactive storytelling [12], we rely on emergent properties to give the story its structure, feeling and style. Although other genres, such as role-playing games, also justify their narrative structure on emergent properties, we believe our approach is somewhat more justified on this reliance. Many of the expected difficulties of controlling story structure have been eradicated because humans drive our story knowledge base and writing process. However, we concede that most authors on weblogs are not literary masters and the degree to which good narrative theory is available in our corpus is limited.

2.1 Story Corpus

To give *Say Anything* its generative power, a large corpus of stories is required. For this work we used the weblog story corpus developed by Gordon et al. [13]. The amount of user-generated content on the Web is rapidly growing and it is estimated that over 70 million weblogs exist [14]. Although written by ordinary people, not all weblog content is relevant to our goal of acquiring narrative content. Gordon et al. [13] estimated that only 17% of weblog writing is actually story content (descriptions and interpretations of causally related past events). The remaining text consists of political and other commentary, lists, opinions, quotations and other non-narrative subject matter. In their work they developed several automatic approaches for extracting the story text from these weblogs that incorporated bag-of-words, part-of-speech, kernel filtering methods and machine learning techniques. Since complete stories were more valuable than excluding non-story content, a version of their system that favored recall over precision was applied to a corpus of 3.4 million weblog entries. This identified 1.06 billion words of story content. Some post processing was necessary in order to sentence delimit the text and remove sentence fragments that were created as a result of the extraction process. The resulting story collection consisted of 3.7 million story segments for a total of 66.5 million sentences.

2.2 Sentence Retrieval

The fundamental mechanism of *Say Anything* is the identification and retrieval of similar stories in our database to the user's input. Contemporary information retrieval techniques provide exactly what we are looking for. In this work we used Apache Lucene [15], a freely available open source search engine. Lucene's algorithm combines standard Boolean indexing with term frequency-inverse document frequency (TF-IDF) scoring functionality. An index is basically a lookup table whose keys are all the unique words/tokens contained in the document collection and the value is the list of all documents that contain that word. Various algorithms exist that allow you to query this index using Boolean operators efficiently. Given the set of documents these words appear in, TF-IDF is a common strategy for scoring and ranking them. All things being equal, words appearing frequently in a document should be given more significance. However, all things are not equal, and words that appear in many documents are given less weight because they are unlikely to be distinguishing words.

To implement this phase of the system, we merely need to treat each of the 66.5 million sentences as a document to be indexed. Simple tokenization is applied and each word is treated as a token in a bag-of-words approach. We then find the highest ranked sentence in our database and return the next sentence of that story.

2.3 The Interface

To complete the system, we needed a user interface. For this project we felt that a web-based interface would be most appropriate since the specifications were rather simple and no special graphics technology was needed. Going this route also allows good cross-platform compatibility and precludes the user from having to install special software on his or her machine.

The interface is split into two main regions. At the top is a panel for viewing the story, which is initially empty. At the bottom are controls for writing and appending sentences to the story. The process starts with the user writing a sentence to begin the story. Given what the user has typed, the system returns the next sentence in the emerging story. When the user feels the story has reached a good concluding point or the story has deviated too far off track (which is often the case), the user may end the story. At this point the user is asked to rate the story on two criteria using a 5 point scale:

1. *Coherence*: Do the system-generated sentences follow from what the user has written? Is the story coherent as a whole?
2. *Entertainment*: Did you have fun with the system? Was the story interesting or entertaining?

In order to successfully complete the story, the user must assign it a title.

3 Analysis

For our initial evaluation we set up a server on our lab's internal network to run our system. We recruited 11 people who wrote a total of 27 stories with the system over a period of two weeks. Our population consisted of men and women who worked in and around our lab, as well as a few other interested parties. We gave the users very little instructions except that they must write and rate their story with the system. We believed at this stage of our research that we should not limit the users to our preconceived notion of how the system should be used; moreover we hoped creative story writers could find functionality beyond our intentions.

There was a wide variety in story generation from the various users. The shortest story consisted of only 2 sentences (1 turn), the longest 30 sentences and an average of 12.7. On average the users had a positive experience using the system 3.39 (1.23 stdev) on a rating scale from 1 (low) to 5 (high). Generally the users did not believe the stories were very coherent. The average coherence rating was 2.29 (1.24 stdev) also on a 5 point scale. See Fig. 1 for a more complete summary of the ratings.

3.1 When Things Go Well

Although the ratings give some indication about the quality of the system, the small sample size and inability to get a feel for what was written limited our knowledge about the system's capabilities. It is more informative to examine some sample output. Fig. 2 presents a complete story from the collection. In some cases it is striking how close the user's sentence matches one from the database (for example those in the first two turns). Despite the poor matching in the third turn, the story does not lose continuity until the fourth turn. Even here, however, the user is able to make the story fairly coherent by shifting the setting to the police station. Although there are some minor peculiarities, the story remains on track until the end. Without any explicit model underneath to guide the process, this story is able to produce a fairly regular narrative structure—an introduction, scene setting, plot development and a climax. The final narrative element, denouement, occurs with the character realizing

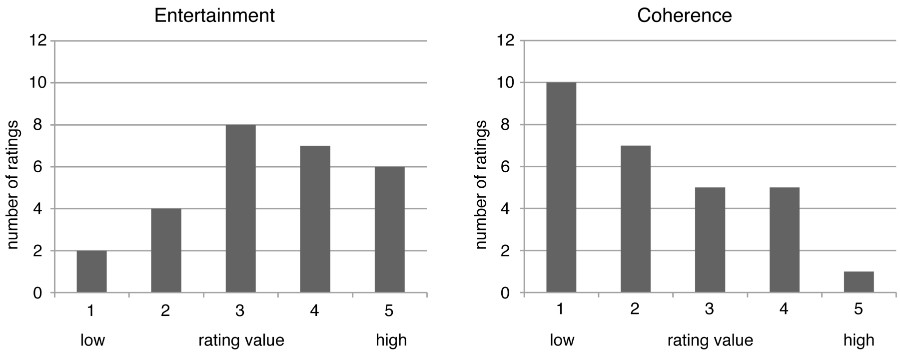

Fig. 1. The number of ratings 11 users gave on 27 stories for the entertainment of the system (left) and the coherence of their story (right).

car chase		
The user's sentence	**Matching sentence**	**Returned sentence**
On one sunny afternoon, I was driving down the highway.	cut to: so i was driving down the 110-S, one sunny L.A. december afternoon- i don't remember where i was going, but i was on the freeway in traffic.	and then i heard it.
An police car in my rear view mirror with its siren on full blast	Then he looked in his rear view mirror and saw a Police Patrol Car behind him, blue lights flashing and siren blaring.	I can get away from him with no problem thought the man and he tromped it some more and flew down the road at over 100 mph.
but his car was no match for the souped police interceptor.	He had his mother's car, a souped-up Ford Mustang.	At speeds topping 100 mph, he made for the airport.
But that proved out to be the wrong thing to do.	2. Thing that makes you feel great is: knowing that people that said i couldn't do it, are getting proved wrong 1.	I filled out 200 questions because: waitin on fatty to get here so i can leave!
yes I was at the police station witing for my lawyer to bail me out.	I thought it was hilarious when David showed up at the police station with the money to bail me out of jail.	Ha!
It is never a good idea to run from the law, atleast not in a 100hp car.	Last run in with the Law: never.	knock on wood 11.

Fig. 2. An example of a story, titled *car chase*, created with the system and given a coherence rating of 4. On the left are the sentences the user submitted to the system. In the middle are the highest ranked sentences in the database and on the right are the next sentences returned to the user by the system.

the errors of his ways. This example also shows that the process is fairly robust to spelling errors.

3.2 When Things Do Not Go As Well

The previous example gives a feel for the kinds of stories that can be produced and how the system works, however, there are several issues that have not been highlighted. An excerpt from a different story, illustrated in Fig. 3, reveals several of the prominent problems with the system.

As is the case for most of the stories, the two major problems are coreference resolution and event ordering/prediction. It is rather impressive that the coreference between entities in *car chase* can be interpreted correctly throughout the story. This is remarkable because at this stage of development no special care has been devoted to them. However, as seen by the example in Fig. 3, this is not always the case. The first sentence in this sequence sets the stage for a driving event that takes place in summer. For coherence, the verbs subscripted with the number 1 would make most sense if they were all part of the main driving event, or related to it in some way. However, in the matched sentence the action *drinking* (wine) is completely unrelated. This is worrisome because it increases the likelihood that the following sentence from that story will be unrelated to events that often occur while driving. Fortunately the next sentence, the one returned to the user, is generic enough that it does not completely spoil the coherency. It is not until the character is sledding that repairing the story becomes nearly impossible. Similarly the phrases subscripted with the number 2 should all refer to the same object, namely the *happy pills*. However, since the retrieval strategy only considers one sentence, there is no way during this simple query to unify the objects. In this case it causes the pills to change into mittens and a sled. This also seems to be the reason that the main event, driving, shifts to become a *sledding* event. This example also illustrates how difficult natural language can be when dealing with implied or implicit knowledge. This can be seen with the explicit reference to the phrase *middle of summer* as indicated by the subscript number 3. Although it is not explicitly indicated, there is an implied reference to winter later on when the main character is sledding. This is indicated in the illustration by an implicit phrase [*in winter*], which does not appear in the actual story. A creative writer could probably work his or her way around this change, although two major shifts in expectation in such a short duration are a lot to overcome.

In addition to these issues, several other problems also frequently arise. Similar to the coreference problem is that of number, gender and case agreement. Some problems also arise specifically out of the nature of the ranking algorithm. Since TF-IDF only considers the frequency of terms, some unfortunate side effects occur. For example, without any preprocessing, numbers and other rare tokens pose a significant problem as seen by the input *"She lived in appartment 4311. I spend many pleasant nights there."* entered by the user[1]. There are many things one would hope the system would pick up on to choose the next sentence. However, what the TF-IDF algorithm deemed the most similar was the number *4311*. This is most likely due to the fact only

The user's sentence	Matching sentence	Returned sentence
Anyway I should know better than to *drive*$_1$ through red flag areas in the *middle of summer*$_3$.	I know better than to *drink*$_1$ red wine after ten, but I did it anyway.	I found myself wide away at four AM with little pain in my face, so I took two of my *happy feel good pills*$_2$.
I wished I had *some*$_2$ right then.	I wish I had *a sled and some mittens*$_2$.	I would have been out there with the kids *sledding*$_1$ anyway [*in the winter*]$_3$.

Fig. 3. Two turns from a story that show problems that can easily lead stories of their path. Note that [*in winter*] does not actually appear in the user's story.

[1] This user broke the rules slightly by inputting more than one sentence at a time.

one sentence in the database contained this number giving it a very high inverse document frequency score. However, this token is only rare by coincidence and not because it is meaningful in any way. Modifying words and phrases also significantly impacts the ability for the retrieval to stay on track. For example, negation, statements of belief or qualifying statements have significant semantic importance but often do not play an important role in finding an appropriate matching sentence.

Although some of the problems seen above occur directly from the IR techniques at the sentence level, the much bigger issue is how to ensure that the global features of the story remain consistent throughout. As demonstrated above, the most difficult issues appear to be event sequence compatibility and ensuring named entities refer to the things they are supposed to. This is not an easy problem to solve but solutions may include event prediction and modeling [16][17][18] and integrating coreference resolution techniques [19].

4 Discussion

Achieving an open domain system that is capable of producing high-quality stories is an extremely difficult undertaking. One of the common criticisms of traditional interactive storytelling mechanisms is the limited domain in which the user can interact with them. Even when the domain is heavily restricted, the cost associated with authoring the content is considerably expensive. For example, Façade [20] one of the premier systems today took two man years to author the content. Although the resulting content and system are among the best, the simulation is only playable for 20 minutes 6 or 7 times. The breadth of coverage and rapid development time are highly attractive despite there being a long way to go before our system produces significant, well structured stories.

Although similar in many respects, our system differs from prototypical interactive storytelling systems in a few key ways. Our goal is not to guide agents in a virtual world in order to tell a pre-authored story, but instead to provide a virtual collaborative writing environment in which the human and computer take turns authoring an entirely new story. Also most interactive story systems take a top down approach where high-level knowledge is hand encoded. In our system we take more of a bottom-up approach where knowledge is acquired from a large corpus of existing stories. This approach allows us to scale to any domain covered in our database and removes many of the difficult narrative theory concerns from the system architecture. In a traditional approach where individual experts write the content for each domain, scaling to the breadth of topics that are common in every day life will be prohibitively expensive. Instead, our system leverages the massive amounts of users who provide this knowledge, not in any logical form, but in plain English that takes no special skill to contribute.

5 Conclusions

In this paper we have presented a new type of interactive storytelling architecture that is unique both in its user interaction model and in the way the system models the story generation process. The sample output and coherence ratings show that there is still a long way to go before high quality stories are capable of being produced. However, there are several encouraging factors that underscore the promise of this approach. One of the most attractive aspects of this architecture is the time of content development. On the one hand the number of man-years required to author all of the content is enormous, consisting of tens-of-millions of weblog users who continue to provide more data than can be reasonably integrated into this system. Although the stories meander without a strong direction more often than not, this methodology shows it is possible that a decent narrative structure can be accomplished using a data-driven model, especially with continued refinement.

Acknowledgements. The project or effort described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

References

1. Crowther, W., Woods, D.: Adventure. (1977)
2. Bates, J.: The nature of characters in interactive worlds and the oz project. Technical report, Carnegie Mellon University, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA (October 1992)
3. Cavazza, M., Lugrin, J.L., Pizzi, D., Charles, F.: Madame bovary on the holodeck: immersive interactive storytelling. In: MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia, New York, NY, USA, ACM (2007) 651–660
4. Riedl, M.O., Young, R.M.: An objective character believability evaluation procedure for multi-agent story generation systems. (2005) 278–291
5. Swartout, W., Gratch, J., Hill, R.W., Hovy, E., Marsella, S., Rickel, J., Traum, D.: Toward virtual humans. AI Mag. **27**(2) (2006) 96–108
6. Bates, J.: Virtual reality, art, and entertainment. Presence: Teleoper. Virtual Environ. **1**(1) (1992) 133–138
7. Montfort, N.: Generating narrative variation in interactive fiction. PhD thesis, Philadelphia, PA, USA (2007) Adviser-Mitchell P. Marcus and Adviser-Gerald Prince.
8. Dungeons & Dragons official website, http://www.wizards.com/ (2008)
9. Wikinovels, http://bluwiki.com/go/Wikinovels (2008)
10. Conklin, J.: Hypertext: An introduction and survey. Computer **20**(9) (1987) 17–41
11. Rettberg, S.: All together now: Collective knowledge, collective narratives, and architectures of participation. In: Digital Arts and Culture. (2005)
12. Cavazza, M., Pizzi, D.: Narratology for interactive storytelling: A critical introduction. In: Technologies for Interactive Digital Storytelling and Entertainment. (2006) 72–83

13. Gordon, A.S., Cao, Q., Swanson, R.: Automated story capture from internet weblogs. In: K-CAP '07: Proceedings of the 4th international conference on Knowledge capture, New York, NY, USA, ACM (2007) 167–168
14. Technorati State of the Blogosphere / State of the Live Web. http://www.sifry.com/stateoftheliveweb (2007)
15. Hatcher, E., Gospodnetic, O.: Lucene in action. Paperback (2004)
16. Bejan, C.A.: Unsupervised discovery of event scenarios from texts. In: The Florida Artificial Intelligence Research Society, Coconut Grove, FL (May 2008)
17. Manshadi, M., Swanson, R., Gordon, A.S.: Learning a probabilistic model of event sequences from internet weblog stories. In: The Florida Artificial Intelligence Research Society, Coconut Grove, FL (May 2008)
18. Chambers, N., Jurafsky, D.: Unsupervised learning of narrative event chains. In: Proceedings of ACL-08: HLT, Columbus, Ohio, Association for Computational Linguistics (June 2008) 789–797
19. Elsner, M., Charniak, E.: Coreference-inspired coherence modeling. In: Proceedings of ACL-08: HLT, Short Papers, Columbus, Ohio, Association for Computational Linguistics (June 2008) 41–44
20. Mateas, M., Stern, A.: Façade: An experiment in building a fully-realized interactive drama. In: Game Developer's Conference, San Jose, California (March 2003)