

# Narrative Intelligence Without (Domain) Boundaries

Boyang Li

School of Interactive Computing, Georgia Institute of Technology, Atlanta, Georgia, USA  
boyangli@gatech.edu

## Abstract

Narrative Intelligence (NI) can help computational systems interact with users, such as through story generation, interactive narratives, and believable virtual characters. However, existing NI techniques generally require manually coded domain knowledge, restricting their scalability. An approach that intelligently, automatically and economically acquires script-like knowledge in any domain with strategic crowdsourcing will ease this bottleneck and broaden the application territory of narrative intelligence. This doctoral consortium paper defines the research problem, describes its significance, proposes a feasible research plan towards a Ph.D. dissertation, and reports on its current progress.

## Introduction

Cognitive research suggests that narrative is a cognitive tool for situated understanding (Bruner 1991). Computational systems that can understand and produce narratives are considered to possess Narrative Intelligence (NI) (Mateas and Sengers 1999). These systems may be able to interact with human users naturally because they understand collaborative contexts as emerging narratives and are able to express themselves by telling stories.

To reason about the vast possibilities a narrative can encompass, a competent NI system requires extensive knowledge of the world where the narratives occur, such as laws of physics, social mechanism and conventions, and folk psychology. Existing systems utilize different forms of knowledge, ranging from small building blocks like actions with preconditions and effects that constitute a bigger story (e.g. Meehan 1976; Riedl and Young 2010; Porteous and Cavazza 2009), to complex structures such as known story cases or plans to be reused or adapted (e.g. Gervás et al. 2005; Li and Riedl 2010).

Obtaining the needed knowledge remains an open problem for narrative intelligence research.

Traditionally, knowledge structures are manually designed for small domains for demonstration purposes. Even when algorithms are general, operating in a new domain would require additional authoring effort, which limits the scalability of NI systems. On the other hand, systems that attempt to mine such knowledge from large text corpora completely automatically (e.g. Swanson and Gordon 2008; Chambers and Jurafsky 2009; McIntyre and Lapata 2010) are limited by the state of the art of natural language processing (NLP) technology.

Between total reliance on human and total automation, this paper investigates a middle road. I propose that by strategically utilizing and actively seeking human help in the form of crowdsourcing, an NI system can acquire the needed knowledge just in time, accurately, and economically. We believe the ability to automatically acquire knowledge about new domains in a just-in-time fashion can lead to *practical* narrative intelligence can generate stories about any given topic, and virtual characters that behave naturally in any setting.

## Research Problem

For applications transcending domain boundaries, a flexible representation is needed that can facilitate powerful story generation algorithms. The *Script* representation (Schank and Abelson 1977), and its variants, is a promising choice. Scripts describe events typically expected to happen in common situations. They provide a default case that story understanding can base its inference on in absence of contradicting evidence. To create a story, one may selectively execute the scripts (Weyhrauch 1997), or break these expectations to create surprising and novel stories (Iwata 2008). Scripts also lend believability to background virtual characters that follow them.

I propose the following requirements (CORPUS) for the automatic acquisition of scripts:

1. **Capture situational variations.** The acquired scripts should contain variations and alternatives to situations.
2. **cOst-effective.** The time and cost of knowledge acquisition should be minimized by automating the construction of scripts as much as possible.
3. **Robust.** The process should withstand inputs that contain errors, noise, ambiguous language, and irrelevant events. Robustness to different parameters settings is also desired.
4. **Proactive.** The system should be able to recognize when its own knowledge is insufficient and actively seek to fill the knowledge gap.
5. **User friendly.** The system should not require significant effort or special expertise from its user.
6. **Suitable for computation.** Knowledge structures extracted should allow believable and interesting stories to be easily generated. The entire process also should work reliably with current technologies without assuming, for example, human-level NLP.

We find crowdsourcing to be particularly suitable to meet these requirements. Crowdsourcing is the practice of delegating a complex task to numerous anonymous workers online, which is often capable of reducing cost significantly. A complex task may be split into a number of simpler tasks that are easy to solve for humans without formal training. The solutions are aggregated into a final solution to the original problem. In this work, a complex knowledge representation about a specified situation is acquired by aggregating short, step-by-step narratives written in plain English into a script. Writing such narratives is a natural means for the crowd workers to convey complex, tacit knowledge.

Compared to acquisition from general-purpose corpora, crowdsourcing has the additional advantages of tapping human creativity and reducing irrelevant information. On one hand, there are many situations—such as taking a date to the movies—that do not often occur in existing corpora. On the other, the corpus we collect from the crowd is expected to contain less irrelevant events compared to, for example, a corpus of news articles.

After crowdsourced narratives are collected, they are automatically generalized into a script, i.e. a graph that contains events and their relations, such as temporal ordering or mutual exclusivity. Reasonable and believable stories can be generated by adhering to these relations. In order to confirm uncertain inferences about script structure or provide missing

information, direct and purposeful questions to the crowd should be employed strategically; we believe this can yield better answers compared to merely increasing the corpus size. Actively learning should also help us identify noise and improve robustness.

## Research Plan

A number of challenges must be overcome to realize NI systems that can generate stories or interactive narratives about any given topic. This section enumerates the challenges and proposes solutions.

The first challenge is **crowd control**. Crowd workers are instructed to write a narrative in English describing how a specified situation (e.g. visiting restaurants, going on a date, or robbing banks) unfolds. This early stage is required to ensure the collected narratives are of high quality and are suitable for subsequent processing. This requires a set of clearly stated, easy-to-understand instructions, a friendly user interface, as well as active checking of user inputs. Further, we put restrictions on the language workers can use in order to alleviate the NLP problem. For example, reference resolution is a problem with no reliable solution. The workers are thus instructed to avoid pro-nouns. We also require each sentence to contain only one verb and describe only one event. Thus, the sentences are expected to segment the situation naturally. However, human workers are error prone and may misunderstand, overlook or forget the instructions. A proper data collection protocol is key for user friendliness and data quality. In addition, intelligent algorithms may identify and reward high-achieving workers and exclude responses from careless workers.

The next challenge is to **identify the primitive events**. In contrast to the Restaurant Game (Orkin and Roy 2009), which relies on a known set of actions, our system does not assume a domain model and learns primitive events from scratch. As workers can describe the event using different words, we need to identify sentences that are similar and cluster them into one event. As user inputs may contain irrelevant or infrequent events, the clustering algorithm should be robust under noise and different cluster sizes.

Even with the help from the crowd, identifying semantically similar sentences can be a major challenge. We will draw from the literature of paraphrasing and predicate parsing to deal with this problem. Fortunately, we can rely on the crowd again for correcting mistakes in automated clustering. An interesting research problem is learning from the mistakes pointed out by the crowd. For example, the

verb is usually instructive of the action taken by a character. However, for the "waiting in line" action, the object being waited for contains more information. When the crowd point out that "waiting in line for popcorn" and "waiting in line for movie tickets" represent two distinct events, the system should be able to learn not to make the same mistake again.

The third challenge is to build the actual script knowledge structures, such as that shown in Figure 1, by **identifying relationships between events**. Traditionally, events are ordered by their temporal relations and support each other with causal relations. I have found a majority vote with global thresholds combined with local relaxation works well for identification of temporal relations. Causal relations tend to be difficult to identify, as people habitually omit overly obvious causers of events when telling a story. However, identifying mutual exclusion relations between events to an extent can compensate for this. The combined use of temporal orderings and mutual exclusions identifies events optional to the script. The next section contains more details.

The fifth challenge is the **generation of stories and interactive narratives** based on the script obtained. Stories can be generated following the temporal relations and mutual exclusions between events, as detailed by the next section. In addition, we may generate stories with unusual and interesting events, such as stories about getting kissing on a date, as in Figure 1, based on frequency. It is also possible to generate textual realizations of a story based on natural language descriptions of events provided by the crowd.

The final challenge is **evaluation**. Story generation systems are known to be difficult to evaluate; we plan to evaluate each component of the system before it is evaluated as a whole. Primitive events learned can be evaluated using metrics designed for clustering, such as purity, precision and recall. A script can be evaluated with respect to its coverage of input narratives, and compared to human-authored scripts based on stories they can generate. The entire system will be evaluated based on the variety and quality of stories it can generate.

### Contributions and Limitations

This section enumerates proposed contributions of this research plan. First, we propose an innovative use of crowdsourcing as a means of rapidly acquiring knowledge about a priori unknown domains (though at this stage we assume we know a few major characters and their roles in the domain). The way we use crowdsourcing yields access to human experiences

that may not exist anywhere else and also allows us to make NLP tractable. Second, we have proposed a unique script representation that caters to both the knowledge acquisition and story generation. Unlike traditionally used plan operator libraries, case libraries, or plot graphs, we forgo causal relations, which are difficult to acquire, and instead resort to temporal relations and mutual exclusion relations. With these relations, we can generate a large variety of coherent stories without a comprehensive understanding of causal necessity.

With our script representations, we intend to demonstrate the complete process starting with human interaction and aggregation of worker responses to the generation of stories using the created knowledge structure. An end-to-end solution can help us avoid making unrealistic assumptions. It has been argued that artificial intelligence research—in order to focus on one research topic—sometimes make too many assumptions that may be later found to be unrealistic or to hinder integration. For instance, separating knowledge acquisition from story generation may lead to authoring bottlenecks.

One limitation to note is that we currently learn scripts best from situations that are commonly experienced by people directly or from novels and movies. Thus, we focus on accurately producing stories about common situations. Collecting a story corpus about rare events, such as submarine accidents, may require more sophisticated interactions with human experts and more sophisticated algorithms for dealing with smaller data sets than those explored in our investigation. Also, we focus on generating stories based on the text medium rather than movies, graphic novels, or other forms of media.

## Research Progress and Results

This section reports the current progress. I have attempted three domains: visiting fast food restaurants, movie dates and bank robbery; 32, 63, and 60 narratives have been crowdsourced on Amazon Mechanical Turk for each domain respectively. Figure 1 shows the script learned from the gold standard events in the movie date domain.

In an iterative trial-and-error process, instructions given to the crowd workers have been carefully tuned. We supply a few major characters and their roles in the social situation we are interested in, and ask crowd workers to describe events happening immediately before, during and immediately after the social situation. When a user make an unexpected error, we insert a new instruction to alert users to avoid it, while

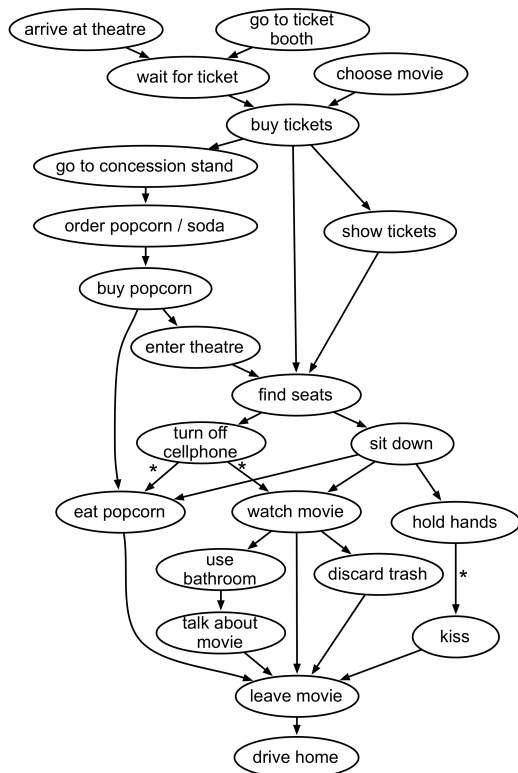


Figure 1. A movie date script. Links denote temporal orderings and asterisks denote orderings recovered by local relaxation.

keeping the instructions as easy to read as possible. The instructions have undergone several iterations and are now relatively stable.

A preliminary study on the identification of primitive events is reported in (Li et al. 2012a, 2012b). Since then, different clustering algorithms and different measures of semantic similarity have been experimented with. At this stage, I found a precision of up to 87% and F1 up to 78%. The next step is to improve the similarity measure by learning from clustering mistakes indicated by additional rounds of crowdsourcing.

A robust method to identify the most common temporal orderings from the narratives was reported in (Li et al. 2012a). Event A is considered to precede event B only when (1) sentences describing A precede sentences describing B much more often than the other way around, and (2) sentences describing A and those describing B co-occur in sufficient number of narratives. These two criteria are maintained by two global thresholds. When crowd workers omit events, or clustering yields imperfect results, significant ordering may fall short of these thresholds. To compensate for this problem, we optimize the on-graph distance between events to produce a graph that better resembles the collected narratives. This

technique effectively relaxes the threshold locally, which leads to robust results under noisy inputs and different parameter settings. It is shown that this technique reduces average error in our graphical structure by 42% to 47%. The asterisks in Figure 1 indicate event orderings recovered by the optimization technique.

Mutual exclusion between events is computed using mutual information, as reported in (Li et al. 2012c). Mutual information measures the interdependence of two events. A mutual exclusion relation is recognized when the occurrences of two events are inversely interdependent and the overall interdependence exceeds a threshold. By combining mutual exclusion with temporal orderings previously identified, we can also identify events that are optional to the script. An algorithm that creates an interactive narrative based on the identified temporal orderings, mutual exclusion relations, and optional events is also reported (Li et al. 2012c). As determined by a brute-force search, the algorithm can generate at least 149,148 unique linear experiences by leveraging the bank robbery data set containing only 60 stories. This demonstrates a good payoff for relatively little authoring cost.

This work is an important first step toward the goal of creating AI systems that minimize the cost of the authoring for story generation and interactive narratives systems. It is the author's vision that reduced authoring cost may one day bring about large-scale applications of AI techniques previously considered to be unscalable.

## References

- Bruner, J. 1991. The narrative construction of reality. *Critical Inquiry* (18):1-21.
- Chambers, N., and Jurafsky, D. 2009. Unsupervised learning of narrative event chains. In *Proceedings of ACL/HLT 2009*.
- Gerrig, R. 1993. *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. Yale University Press.
- Gervás, P., Díaz-agudo, B., Peinado, F., and Hervás, R. 2005. Story plot generation based on CBR. *Knowledge-Based Systems* 18 (4-5):235-242.
- Iwata, Y. 2008. *Creating Suspense and Surprise in Short Literary Fiction: A stylistic and narratological approach*. Ph.D. Dissertation, University of Birmingham.
- Li, B., Appling, D. S., Lee-Urban, S., and Riedl, M. O. 2012a. Learning Sociocultural Knowledge via Crowdsourced Examples. In *the 4th AAAI Workshop on Human Computation*.
- Li, B., Lee-Urban, S., Appling, D. S., and Riedl, M. O. 2012b. Automatically Learning to Tell Stories about Social Situations from the Crowd. In *the LREC 2012 Workshop on Computational Models of Narrative*, Istanbul, Turkey.
- Li, B., Lee-Urban, and Riedl, M. O. 2012c. Toward

- Autonomous Crowd-Powered Creation of Interactive Narratives. In *the INT5 Workshop*, Palo Alto, CA.
- Li, B., and Riedl, M. O. 2010. An offline planning approach to game plotline adaptation. In *AIIDE 2010*.
- Mateas, M., and Sengers, P. 1999. Narrative intelligence. In *Proceedings of the 1999 AAAI Fall Symposium on Narrative Intelligence*. Menlo Park, CA.
- Meehan, J. R. 1976. *The Metanovel: Writing stories by computer*. Ph. D. Dissertation, Yale University.
- McIntyre, N., and Lapata, M. 2010. Plot Induction and Evolutionary Search for Story Generation. In *ACL 2010*, Uppsala, Sweden.
- Orkin, J., and Roy, D. 2009. Automatic learning and generation of social behavior from collective human gameplay. In *AAMAS 2009*, Budapest, Hungary.
- Porteous, P., and Cavazza, M. 2009. Controlling Narrative Generation with Planning Trajectories: The Role of Constraints. In *ICIDS 2009*, Guimarães, Portugal.
- Riedl, M. O., and Young, R. M. 2010. Narrative Planning: Balancing Plot and Character. *Journal of Artificial Intelligence Research* 39:217-268.
- Schank, R. C., and Abelson, R. 1997. *Scripts, Plans, Goals and Understanding*: Lawrence Erlbaum.
- Swanson, R., and Gordon, A. 2008. Say Anything: a massively collaborative open domain story writing companion. In *ICIDS 2008*, Erfurt, Germany.
- Weyhrauch, P. 1997. *Guiding Interactive Drama*. Ph.D. Dissertation, Carnegie Mellon University.