Proceedings, The Thirteenth AAAI Conference on
Artificial Intelligence and Interactive Digital Entertainment (AIIDE-17)

Towards Expressive Automated Storytelling Systems

David R. Winer
School of Computing
The University of Utah
Salt Lake City, UT
drwiner@cs.utah.edu

Introduction

Narrative intelligence is an interdisciplinary area of research in artificial intelligence leveraging insights from linguistics, cognitive psychology, narratology, and computer science (Schank 1995). This work addresses the problem of generating narrative fiction (e.g., text or film) by using a plan-based language to model the schematic knowledge of 1] storyworld mechanics and 2] communicative plans. The science for merging these two tasks is desirable for generating narrative discourse which has the goal-oriented and hierarchical structure to support expressive storytelling. Integrating scenario generation with communicative reasoning has broad applicability for problem-solving agents which interact with people such as in the context of entertainment and education.

My approach to automated narrative generation borrows from narrative theory which frequently distinguishes *fabula* (i.e., setting and plot events, story) from discourse (the communicative act of storytelling) (Chatman 1980).

- ***Fabula*** Actions at the fabula level are storyworld mechanics. Story generators arose as the first AI planning algorithms were developed, such as TALE-SPIN (Meehan 1977) in which woodland creatures follow plans to satisfy basic needs. State-of-the-art planners solve multi-agent coordination problems such that characters follow domain-independent rules to behave more believably (Riedl and Young 2010; Ware et al. 2014).
- **Discourse** The communication side of generating narrative content (storytelling) involves modeling the beliefs of a hypothetical viewer and conveying meaning through utterances which refer to the plot (Winer et al. 2015; Wu, Young, and Christie 2016), consistent with discourse theories from narratology (Chatman 1980).

One of the predominant strategies for designing automated storytelling systems is to adhere to the natural language generation (NLG) pipeline: start with a set of events or a library of information (fabula), generate discourse for conveying the events such as with paragraph and sentence planning, and last make edits to conform to the structural requirements of the medium (Reiter, Dale, and Feng 2000; Callaway and Lester 2002). Systems which borrow this *fabula-then-discourse* architecture typically take fabula as input and form a storytelling plan around some subset of events (Young et al. 2013).

My research addresses two major **research problems**:

1. The *fabula-then-discourse* strategy is not designed well for tasks requiring coordinated plot and communication such as generating narrative fiction. An input fabula (or one generated in isolation) may be coherent and believable, but it may not have the desired attributes for good storytelling plans (e.g., take into account the pragmatics of communicating) and thus the quality of storytelling may suffer. Unless directed, a character agent is not likely to arrange itself for a camera shot or hold its positions at favorable moments, and these actions are typically planned at the fabula level.
2. Expert knowledge for storytelling is difficult to formalize and hand-code, causing an *authorial bottleneck problem* (Valls-Vargas, Zhu, and Ontanon 2016; Riedl and Sugandh 2008). Despite a long research history of encoding communicative knowledge as actions (Cohen and Perrault 1979; Young and Moore 1994; Jhala and Young 2010), it isn't widely practiced in storytelling and resources are limited. Recent work with information extraction has shown progress for informing narrative models from unannotated narrative text (Chambers and Jurafsky 2008; Goyal, Riloff, and Daumé III 2010; Valls-Vargas, Zhu, and Ontanon 2016) and from crowdsourced examples (Li et al. 2013). However, these models are still impoverished in the kinds of features they use related to communication and scene-structure (e.g., sentiment analysis (Li et al. 2014; Reagan et al. 2016)).

In prior work (Winer and Young 2016) and in work accepted to AIIDE 2017 (Winer and Young 2017), I introduce a novel algorithm architecture (BiPOCL) which addresses **research problem 1** through automated merging of fabula and discourse languages to support generating narratives with coordinated story and discourse features. In theory, the algorithm interleaves story and discourse planning (Winer and Young 2016), rather than generating story first, and in practice the coordinated features are baked together in a pre-computing phase before planning begins (Winer and Young 2017). This architecture supports expressive storytelling at-

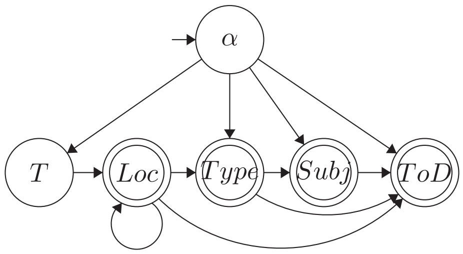

Figure 1: Simplified shot heading automata starting at α. A shot heading is a tuple of the form $\langle T, Loc, ST, Subj, ToD \rangle$ where $T \in \{INT., EXT., INT./EXT., \emptyset\}$, Loc is a list of increasingly specific locations (or empty), ST is a shot type from from an enumerated list of types (or empty), $Subj$ is a word or phrase (or blank), and ToD is a word or phrase for the time of day (or blank).

tributes, but as **research problem 2** suggests, such attributes are manually created as they are difficult to extract.

Screenplays (film scripts) are unusual for narrative text because they contain more structured discourse information (Jhala 2008) than other narrative texts such as news stories (Chambers and Jurafsky 2008) or fables (Goyal, Riloff, and Daumé III 2010; Valls-Vargas, Zhu, and Ontañón 2015) which have received more attention. Screenplays follow a standardized format for their parts (e.g., stage direction, dialogue) including shot headings which include details about what, where, when, and how to film the events in their subsequent section. According to *The Hollywood Standard* (Riley 2009), an authoritative guide to screenplay writing, the shot headings follow a rigid syntax which I've formalized and implemented as part of a screenplay parser (see Figure 1). For example:

```
EXT. WHITE HOUSE - SOUTH LAWN - CLOSE
ON CNN CORRESPONDENT - SUNSET
```

This shot heading indicates that the subsequent stage direction and dialogue is shot outside the White House on the South Lawn at sunset with a close-up. There are many screenplays available online (e.g., the Internet Movie Script Database[1]) potentially providing a wealth of information to inform a model of storytelling for different scenarios in different genres. I've extracted screenplay *segments* from IMSDb, which are shot heading - stage direction pairs and speaker - dialogue pairs. I will next introduce the proposed approach.

Knowledge Extraction

The goal is to extract storytelling patterns from screenplays to address the authorial bottleneck problem referenced in the introduction. The overall goal is to learn hierarchical cinematic narrative discourse patterns for narrative generation (Jhala and Young 2010; Young et al. 2013) where an instantiated pattern would represent a generated segment whose *subplan* consists of actions taken by characters and whose preconditions and effects include conditions associated with features of the segment. These patterns would then be used for *automated cinematic narrative generation*. I focus specifically on segments containing stage direction and leave dialogue for future work. The task of learning hierarchical patterns is broken into stages: 1] *lexical action recognition*, and 2] *schema induction*.

[1] www.imsdb.com

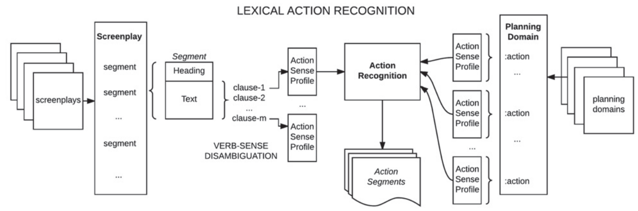

Figure 2: Schematic of the lexical action recognition process

Lexical Action Recognition

After collecting segments as in Figure 2, I propose to map clauses in stage direction to STRIPS-style (Fikes and Nilsson 1972) action schemata for a domain of interest provided as input. The planning domain provided as input may be tailored for a specific genre (e.g., Western shootouts, dragon slaying, etc.) or represent a generic set of action types that can occur in a wide variety of contexts. Action schemata in a planning domain are manually annotated with a set of lexical constraints.

It would not be sufficient to label action schemata with specific verbs because verbs have a variety of meanings and may be part of a multiword phrase (Del Corro, Gemulla, and Weikum 2014). I leverage two widely used lexical databases: FrameNet (Baker, Fillmore, and Lowe 1998) and WordNet (Fellbaum 2010). The FrameNet database uses *frames* which are schematic representations of types of situations such as an action's operation. I use an off-the-shelf parser `Semafor` (Das and Smith 2011) to identify verb and argument frames from an input sentence. Frames are insufficient for the task because they do not commit to a particular instance of a situation (e.g., a *change_position* frame doesn't indicate the direction of change). Thus, I also use the WordNet database which groups words into *synsets*, categories representing synonyms which can be shared among verbs. I use the clause parser `ClausIE` (Del Corro and Gemulla 2013) to prune the set of possibly synsets for a verb instance as done in other work (Del Corro, Gemulla, and Weikum 2014).

The output of the parser is an *action sense profile* which is compared to the lexical constraints on plan actions (see Figure 2).

Definition 1 (Action Sense Profile) *If t is an action or clause, then t's action sense profile is a tuple of the from $\langle S_t, F_t \rangle$ where S_t is a set of synsets associated with the sense of verb uses which can represent t and F_t is a set of frames associated with the intended category for the operation of t.*

Schema Induction

The second stage for learning hierarchical patterns representing screenplay segments is *schema induction*. My methodology will be inspired by similar work such as learning narrative scripts from commonly co-occurring verb instances in news stories (Chambers and Jurafsky 2008) and learning from crowdsourced stories about the same event (Li et al. 2013). The details for this process are still in development and involves the output of the lexical action recognition process mapped to vector representations.

Conclusion

I have developed a generative model which characterizes expressive features of storytelling by integrating fabula and discourse; however, expert knowledge which leverages this model is lacking. I extract fabula and discourse information from screenplays to facilitate storytelling knowledge extraction. The proposed work would benefit future research in narrative understanding and natural language generation by leverage insights from narrative planning.

References

Baker, C. F.; Fillmore, C. J.; and Lowe, J. B. 1998. The berkeley framenet project. In *ACL and COLING*, 86–90.

Callaway, C. B., and Lester, J. C. 2002. Narrative prose generation. *Artificial Intelligence* 139(2):213–252.

Chambers, N., and Jurafsky, D. 2008. Unsupervised learning of narrative event chains. In *ACL*, volume 94305, 789–797.

Chatman, S. B. 1980. *Story and discourse: Narrative structure in fiction and film*. Cornell University Press.

Cohen, P. R., and Perrault, C. R. 1979. Elements of a plan-based theory of speech acts. *Cog. Science* 3(3):177–212.

Das, D., and Smith, N. A. 2011. Semi-supervised frame-semantic parsing for unknown predicates. In *ACL: Human Language Technologies*, 1435–1444. ACL.

Del Corro, L., and Gemulla, R. 2013. Clausie: clause-based open information extraction. In *Proceedings of the 22nd international conference on World Wide Web*, 355–366. ACM.

Del Corro, L.; Gemulla, R.; and Weikum, G. 2014. Werdy: Recognition and disambiguation of verbs and verb phrases with syntactic and semantic pruning. In *ACL*.

Fellbaum, C. 2010. Wordnet. In *Theory and applications of ontology: computer applications*. Springer. 231–243.

Fikes, R. E., and Nilsson, N. J. 1972. Strips: A new approach to the application of theorem proving to problem solving. *Artificial intelligence* 2(3):189–208.

Goyal, A.; Riloff, E.; and Daumé III, H. 2010. Automatically producing plot unit representations for narrative text. In *EMNLP*, 77–86. ACL.

Jhala, A., and Young, R. M. 2010. Cinematic visual discourse: Representation, generation, and evaluation. *TCIAIG* 2(2):69–81.

Jhala, A. 2008. Exploiting structure and conventions of movie scripts for information retrieval and text mining. *Interactive Storytelling* 210–213.

Li, B.; Lee-Urban, S.; Johnston, G.; and Riedl, M. 2013. Story Generation with Crowdsourced Plot Graphs. In *AAAI*.

Li, B.; Thakkar, M.; Wang, Y.; and Riedl, M. O. 2014. Storytelling with adjustable narrator styles and sentiments. In *ICIDS*, 1–12. Springer.

Meehan, J. R. 1977. Tale-spin, an interactive program that writes stories. In *IJCAI*, volume 77, 91–98.

Reagan, A. J.; Mitchell, L.; Kiley, D.; Danforth, C. M.; and Dodds, P. S. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science* 5(1):31.

Reiter, E.; Dale, R.; and Feng, Z. 2000. *Building natural language generation systems*, volume 33. MIT Press.

Riedl, M., and Sugandh, N. 2008. Story planning with vignettes: Toward overcoming the content production bottleneck. *Interactive Storytelling* 168–179.

Riedl, M. O., and Young, R. M. 2010. Narrative planning: Balancing plot and character. *JAIR* 39(1):217–268.

Riley, C. 2009. *The Hollywood standard: the complete and authoritative guide to script format and style*. Michael Wiese Productions.

Schank, R. C. 1995. *Tell me a story: Narrative and intelligence*. Northwestern University Press.

Valls-Vargas, J.; Zhu, J.; and Ontañón, S. 2015. Narrative hermeneutic circle: improving character role identification from natural language text via feedback loops. In *IJCAI*, 2517–2523.

Valls-Vargas, J.; Zhu, J.; and Ontanon, S. 2016. Error analysis in an automated narrative information extraction pipeline. *TCIAIG*.

Ware, S. G.; Young, R. M.; Harrison, B.; and Roberts, D. L. 2014. A computational model of plan-based narrative conflict at the fabula level. *TCIAIG* 6(3):271–288.

Winer, D. R., and Young, R. M. 2016. Discourse-driven narrative generation with bipartite planning. In *INLG*, 11–20.

Winer, D. R., and Young, R. M. 2017. Merits of hierarchical story and discourse planning with merged languages. In *AIIDE 2017*.

Winer, D. R.; Amos-Binks, A. A.; Barot, C.; and Young, R. M. 2015. Good timing for computational models of narrative discourse. In *CMN Workshop*, volume 45.

Wu, H.-Y.; Young, M.; and Christie, M. 2016. A cognitive-based model of flashbacks for computational narratives. In *AIIDE*.

Young, R. M., and Moore, J. D. 1994. DPOCL: A principled approach to discourse planning. In *INLG*, 13–20. ACL.

Young, R. M.; Ware, S.; Cassell, B.; and Robertson, J. 2013. Plans and planning in narrative generation: a review of plan-based approaches to the generation of story, discourse and interactivity in narratives. *Sprache und Datenverarbeitung, Special Issue on Formal and Computational Models of Narrative* 37(1-2):41–64.