Computational Models of Narrative: Review of the Workshop

Mark A. Finlayson, Whitman Richards, and Patrick H. Winston

■ *On October 8–10, 2009, an interdisciplinary group met in Beverley, Massachusetts, to evaluate the state of the art in the computational modeling of narrative. Three important findings emerged: (1) current work in computational modeling is described by three different levels of representation; (2) there is a paucity of studies at the highest, most abstract level aimed at inferring the meaning or message of the narrative; and (3) there is a need to establish a standard data bank of annotated narratives, analogous to the Penn Treebank.*

Narratives are ubiquitous in human experience. We use them to entertain, communicate, convince, and explain. One workshop participant noted that "as far as I know, every society in the world has stories, which suggests they have a psychological basis, that stories do something for you." To truly understand and explain human intelligence, reasoning, and beliefs, we need to understand why narrative is universal and explain the function it serves.

Computational modeling is a natural method for investigating narrative. As a complex cognitive phenomenon, narrative touches on many areas that have traditionally been of interest to artificial intelligence researchers: its different facets draw on our capacities for natural language understanding and generation, commonsense reasoning, analogical reasoning, planning, physical perception (through imagination), and social cognition. Successful modeling will undoubtedly require researchers from these many perspectives and more, using a multitude of different techniques from the AI toolkit, ranging from, for example, detailed symbolic knowledge representation to large-scale statistical analyses. The relevance of AI to narrative, and vice versa, is compelling.

The Computational Models of Narrative workshop[1] had three main objectives: (1) to understand the scope and dimensions of narrative models, identifying gaps and next steps, (2) to evaluate the state of the art, and (3) to begin to build a community focused on computational narrative. The interdisciplinary group of 22 participants (see figure 1) included computer scien-

tists, psychologists, linguists, media developers, philosophers, and storytellers. Ten speakers were selected to represent a range of views, and their presentations were organized into four groups, each followed by an extensive discussion moderated by a panel. The day after the presentations, there was a lively, morning-long extended discussion. The meeting's audio was captured and later analyzed in depth. A detailed summary of the group's conclusions at the workshop appears elsewhere (Richards, Finlayson, and Winston 2009), together with recommendations for future initiatives.[2] Regarding models of narrative, the main findings were: (1) a three-level organization of narrative representations unifies work in the area, (2) the area suffers from a deficit of investigation at the highest, most abstract level aimed at the "meaning" of the narrative, and (3) there is a need to establish a standard data bank of annotated narratives, analogous to the Penn Treebank (Marcus, Marcinkiewicz, and Santorini 1993).

A Three-Level Organization

Computational modeling requires a precise statement of the problem (or problems) to be solved. Thus, an obvious first step is to understand how narrative should be represented.

There were three common denominators among the representations presented at the workshop: (1) narratives have to do with sequences of events, (2) narratives have hierarchical structure, and (3) they are grounded in a commonsense knowledge of the world. Similarly, it was uncontroversial that narratives can be told from multiple points of view, and that all four of these characteristics were independent of whether or not a narrative was told with words.[3]

After analysis of the presentations and discussions, it became clear that all the representations considered at the workshop were subsumed within a three-level structure. The heavily investigated middle level stressed event sequences that were built on the classic logical-predicate-like representations introduced in artificial intelligence in its earliest days, exemplified by instances such as KISS(JOHN, MARY) and CAUSE(SHOOT, DIE).

Below the middle level were representations that examined the detailed structure of the narratives in question. There was quite a bit of work at this detail level, such as commonsense reasoning (Mueller 2007), discourse structures (Asher and Lascarodes 2003), argument-support hierarchies (Bex, Prakken, and Verheij 2007), or plan graphs (Young 2007).

Above the middle level was a third, more abstract kind of representation. This abstract level encoded structures that were not directly present in the story itself but had to be inferred in light of a larger context. Structures include plot functions (in Vladimir Propp's sense), plot units (Wendy Lehnert) or the simply stated "meaning" of a narrative. Only a single piece of work presented during the workshop dealt with this upper level (Finlayson 2009), and it became clear during the discussion that there was a paucity of work at this level. Yet all participants agreed that tackling this level was crucial to truly understanding the character of narrative.

Questions

In the course of discussions, three questions were raised again and again, and it was clear that participants were far from consensus on answers to any of them: Why narrative? What are the appropriate representations of narrative? What are appropriate experimental paradigms for narrative?

Why narrative? Participants were unable to give a concise and cogent reason why narrative per se, rather than a variety of other cognitive or social processes, should be an object of study. Several participants claimed narrative was only an epiphenomenon, where the real target of inquiry should be, say, analogy, planning, or social interaction. In any event, all the proposed answers still begged the question of what is special about narrative in particular—there are many ways of structuring events, but not all of them are a narrative. Although numerous examples and dimensions were proposed and discussed, there was no consensus on a definition for narrative, no procedure for distinguishing narratives from nonnarratives, and no procedure for distinguishing good narratives from bad.

What are the appropriate representations of narrative? While there was some shared core structure to the representations presented (noted previously—the middle level), researchers' representations varied widely and covered different parts of the representational spectrum. Numerous "dimensions" of narrative were identified, but there was no consensus as to how the dimensions mapped into or were split across representations.

What are appropriate experimental paradigms for narrative? Related to the above was the issue of how the different representations dictated the form of experimental paradigms needed to validate the computational models. Consistent with uncertainties in precise computational definitions, there was also no consensus on what kinds of experiments would be the most informative. Should model evaluations be based on question answering, or on story simulation, or on analogies between narratives that reflected human judgments?

Figure 1. Workshop Participants in Beverley, Massachusetts, Thursday, 9 October 2009.

Back row, left to right: J. Keyser (MIT), I. Horswill (Northwestern), M. Young (North Carolina State University), B. Verheij (Groningen), M. Cox (DARPA), S. Narayanan (ICSI and Berkeley), T. Lyons (AFOSR), L. Jackson (Naval Postgraduate School); *Middle row, left to right:* H. Lieberman (MIT), K. Forbus (Northwestern), M. Finlayson (MIT), E. Mueller (IBM), P. Winston (MIT), N. Asher (Texas), J. Hobbs (USC ISI), V. Subrahmanian (Maryland); *Front row, left to right:* N. Cohn (Tufts), R. Jackendoff (Tufts), P. Gervás (U. Complutense Madrid), W. Richards (MIT), R. Swanson (USC ICT), E. Tomai (Pan American), M. Seifter (MIT).

Courses of Action

One goal of the workshop was to identify next steps that could further progress in the area. In the intensive discussion on the last day of the workshop, at least three of the courses of immediate action proposed were universally applauded.

First, the participants agreed that the workshop was a boost to understanding narrative, by bringing together a variety of approaches and showing links and differences. The group felt the community was fragmented and needed to be encouraged and grown. Many of the participants had not previously met, and consequently a variety of perspectives and approaches were new to large segments of this small group. A second workshop would be the obvious next step toward establishing a larger, still broader community. We visualize the second workshop as doubling in size and including several areas not represented, such as game-theoretic approaches to narrative, studies of gossip and rumor, and narrative theory researchers from the humanities. It was generally agreed that more thought is also needed to reach an agreement on methods for evaluating story understanding, as well as various experimental paradigms.

In addition to broadening the scope of participants, a second workshop is needed to investigate whether a new community should be set up (with its own annual meetings and publication vehicles) or whether the participants are naturally a subset of an already established community. In particular, some participants laid great stress on investigating whether a publication venue directly associated with the area would be appropriate.

Second, it was suggested a catalogue be assembled listing potential applications of narrative—big problems on which narrative might give traction. An obvious example would be comparing news reports from different perspectives or cultures. Such a catalogue would be of great use to motivating work and securing funding.

Third, nearly every participant noted the sore lack of a shared corpus of stories—a necessary tool if one is to compare successes and strengths of various approaches. Hence it was proposed to create a story databank. One important property of such a collection would be to provide at least one computationally tractable representation of the texts in an agreed-upon format. It was acknowledged that annotation is a time-consuming and delicate

NEW!

AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI)

The first AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI) will be held in conjunction with AAAI-10 in Atlanta. The EAAI symposium provides a venue for AI researchers involved in education to share their innovative approaches to education and teaching. In contrast to work on using AI as a building block in educational systems (such as intelligent tutoring systems), EAAI focuses on pedagogical issues related to teaching AI at a variety of levels (from K–12 through postgraduate training).

The EAAI symposium is comprised of several components, including a program of high-quality refereed papers, panels, special sessions, and invited talks; a presymposium workshop for mentoring new faculty, instructors, and teaching assistants; an Educational and Teaching Video track within the AAAI Video Program; a Student and Educator Robotics track within the AAAI Robotics Exhibition and Workshop; and a poster session, held in conjunction with the AAAI poster session.

For more information about the symposium, please visit the AAAI-10 website or write to us at aaai10@aaai. org.

process, but despite this obstacle there was general agreement that there should be some attempt to make a story databank. A committee was set up to pursue the task as well as to decide upon an annotation format. Interested parties should contact Pablo Gervás at the Universidad Complutense de Madrid or Mark Finlayson at the Massachusetts Institute of Techology.

References

Asher, N., and Lascarides, A. 2003. *Logics of Conversation.* New York: Cambridge University Press.

Bex, F. J.; Prakken, H.; and Verheij, B. 2007. Formalizing Argumentative Story-Based Analysis of Evidence. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law,* 1–10. New York: Association for Computing Machinery.

Finlayson, M. A. 2009. Deriving Narrative Morphologies via Analogical Story Merging. In *New Frontiers in Analogy Research: Proceedings of the Second International Conference on Analogy,* 127–136. Sofia, Bulgaria: New Bulgarian University Press.

Marcus, M. P.; Marcinkiewicz, M. A.; and Santorini, B. 1993. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics* 19(2): 313–330.

Mueller, E. T. 2007. Modeling Space and Time in Narratives about Restaurants. *Literary and Linguistic Computing* 22(1): 67–84.

Richards, W.; Finlayson, M. A.; and Winston, P. H. 2009. Advancing Computational Models of Narrative. CSAIL Technical Report No. 2009-063. Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory, Cambridge, MA.

Young, R. M. 2007. Story and Discourse: A Bipartite Model of Narrative Generation in Virtual Worlds. *Interaction Studies* 8(2): 177–208.

Mark Alan Finlayson is a doctoral candidate in electrical engineering and computer science at the Massachusetts Institute of Technology. His research interests are in computational models of human intelligence, in particular, the effect of culture and narrative on cognition. He holds a B.S. from the University of Michigan and an M.S. from MIT, both in electrical engineering.

Whitman Richards is a professor of cognitive sciences in the Department of Brain and Cognitive Sciences at the Massachusetts Institute of Technology and has been on the MIT faculty since 1965. He is also a professor of cognitive science, media arts, and sciences at the MIT Media Lab. He holds a B.S. in physical metallurgy (materials science) and a Ph.D. in experimental psychology, both from MIT.

Patrick Henry Winston is the Ford Professor of Artificial Intelligence and Computer Science at MIT. He has been on the MIT faculty since 1970 and was director of the MIT Artificial Intelligence Laboratory from 1972 to 1997. He served as president of AAAI from 1985 to 1987. His principle research interest is in how vision, language, and motor faculties together account for intelligence. He holds B.S. and M.S. degrees in electrical engineering and a Ph.D. in computer science, all from MIT.