Logical Abduction as a Computational Model of Narrative

Andrew S. Gordon

University of Southern California, Institute for Creative Technologies, 12015 Waterfront Drive, Los Angeles, CA 90094 USA

Abstract

Abductive reasoning, distinct from deductive and inductive reasoning, has often been characterized as "inference to the best explanation." In this paper, I argue that many aspects of narrative interpretation and generation can be readily modeled in terms of abductive reasoning, and that software algorithms for automated logical abduction offer an attractive foundation for computational models of narrative.

Keywords

Logical Abduction, Narrative Interpretation, Narrative Generation

1. Introduction

In the pursuit of a better scientific understanding of the cognitive abilities that people have for narrative generation and interpretation, computational models of these cognitive mechanisms can serve to validate hypotheses, distinguish theoretical approaches, and uncover previously unseen research questions. These benefits can be had even if the pursuit of these computational models never produces software capable of replicating these human abilities in any meaningful way. Interestingly, the reverse also appears to be true; the remarkable abilities of large language models to generate and seemingly understand complex narratives has not yet advanced our scientific understanding of how the human brain accomplishes these feats. While today it is easy to download and execute a (deep neural network) model with narrative competencies, the advancement of a (computational cognitive) model of narrative in support of scientific understanding remains an import research endeavor.

In this writeup I highlight one specific computational model, based on logical abduction, that has proven useful in furthering our understanding of discourse interpretation over the last several decades. This computational model grew out of symbolic approaches to engineering AI systems for language understanding in the field of computational linguistics, a research field that today has almost entirely moved away from symbolic and logic-based approaches in favor of deep neural networks and large language models. Here I argue that this computational model based on logical abduction provides useful insight into our cognitive abilities for narrative interpretation. Furthermore, I explore the potential for this model to serve as a basis for computational models of the complimentary abilities that people have for narrative generation.

2. Logical Abduction

The philosopher Charles Pierce (1839-1914) described logical abduction as a formal reasoning method, distinct from deduction or induction, that searches for assumptions that, if they were indeed true, would logically entail a set of observations given a knowledge base of logical axioms. In the tradition of formal logic, an *axiom* can be viewed as generalized knowledge about the

CMN 2025 : 8th International Workshop on Computational Models of Narrative, May 28-30, 2025, Geneva, Switzerland.

gordon@ict.usc.edu

world that we accept as always true, and *entailment* can be seen as the process that allows these axioms to derive additional truths about the world using the rules of logic. Abduction runs this process in reverse, looking for hypotheses that may or may not be true which would (logically) account for the observations that are already accepted as true. With this perspective, Pierce offers a way of reasoning about the unobserved and unobservable world based on the observable evidence at hand, providing a logic-based approach to answering *Why* questions.

Already from this short description it may seem that the mechanisms of logical abduction are almost completely orthogonal to concerns of narrative. With its harsh division between what is true or false and its strict requirement for axiomatic world knowledge that is never wrong, formal logic is certainly a tool better suited for describing mathematical proofs than for reasoning about the ambiguities inherent in narratives of the complex lives of people and the intermingling of fictional and nonfictional realities.

The early computational linguistics work of Jerry Hobbs and colleagues [1] provided the insight that helped bridge this conceptual divide. In their proposal for "interpretation as abduction" they saw the cognitive process of language understanding as an abductive reasoning task, where the observable truths are the words that are given, and the mental task is to figure out *Why*. That is, given the common knowledge that readers and writers share, what meaning was a writer trying to impart with the words they have chosen? In logical terms, what unobservable mental representations of meaning would logically entail the discourse that is to be understood?

To serve as a computational model of discourse interpretation, two important difficulties of logical abduction needed to be addressed. First, needed was some way of representing the wealth of knowledge that is shared by the discourse participants as (always true) logical axioms, to include knowledge of language and commonsense background knowledge of the way the world works. For this, Hobbs extended early work by AI-pioneer John McCarthy to propose a way of writing defeasible axioms in formal logic by including special *etcetera* literals in inference rules, i.e., conditions that must be assumed true (via abduction) for the knowledge to be truly axiomatic. Second, needed was an algorithm that used this knowledge to find the best (unobservable) mental representations of meaning for the observable discourse. For this, Hobbs and colleagues proposed Weighted Abduction [1], an algorithm that identified possible meanings that could be found by backtracking from observed words using knowledgebase axioms and ranking them according to weighted costs assigned to each axiom.

My own contributions to the computational aspects of this model came many years later. First, I established a probabilistic formulation of etcetera literals that simplified the ranking of possible interpretations and provided opportunities to learn the probabilities associated with knowledgebase axioms from data [2]. Second, I implemented an efficient algorithm capable of managing the combinatorial search challenge for large interpretation problems, called Incremental Etcetera Abduction [3].

3. Narrative Interpretation

We can see a compelling example of the use of logical abduction as a computational model of narrative discourse in its application to the interpretation of the famous Heider-Simmel film. Created in the 1940s by social psychologist Fritz Heider and his student Marianne Simmel [4], this 90-second animated silent film depicts the 2D motions of two triangles and circle in and around a box with an articulated "door". Despite the simplicity of the animation, the experimental subjects who viewed this film readily perceived the animation as a story involving two characters (the circle and the smaller triangle) being assaulted by a third character (the larger triangle). The circle retreats inside of the box and is cornered by the larger triangle but manages to escape with the smaller triangle by trapping the larger triangle in the box. Later in his influential book *The Psychology of Interpersonal Relations* [5], Heider proposed that his experimental subjects were compelled to anthropomorphize these simple shapes as people with plans, goals, and emotions to explain their on-screen trajectories, employing a commonsense theory of human psychology to understand the narrative of the film.

In previous work, I modeled this narrative interpretation of the Heider-Simmel film using logical abduction, specifically an implementation of Incremental Etcetera Abduction [3]. In this work, the observable truths were a sequence of logical literals encoding 76 recognizable actions of the characters in the film, e.g., the opening of the door by the smaller triangle and the shaking motion of the circle. The knowledge base used to interpret these actions (via backtracking from the observables) consisted of 81 axioms that encoded the commonsense relation between mental states and behavior, e.g., that someone opening a door might do so because they want to go inside, and that someone who is shaking might do so out of fear. Applying the Incremental Etcetera Abduction algorithm with these inputs launches a combinatorial search through all possible explanations of all observations, recursively executed to a specified depth and conducted in incremental batches of observations to contain the size of the search space. The joint probability of each potential interpretation is computed as a product of the assigned probabilities of assumed etcetera literals, and the most-probable solution is selected as the best interpretation.

From the perspective of narratology, the most interesting aspect of logical abduction is the output of this search process, which is a proof graph that connects assumptions about unobservable world states, such as the mental states of the storyline characters, to their observable behaviors. A key aspect of this algorithm for logical abduction is that assumptions that play a role in the explanation of two different observations are combined wherever possible (via the process of first-order unification). This unification of assumptions helps promote interpretations that include common factors, both figuratively and literally, as unifying two assumptions will drop a factor when making the joint probability calculation, increasing its likelihood estimation. As a result, the best (most likely) interpretations (proof graphs) are those that tend to have structures that share a remarkable resemblance to the Story Intention Graphs [6] and Causal Network Models [7] that have been manually encoded in previous structural analysis of narratives and have also proven instrumental in algorithms for automatic generation of narrative text [8]. If graphs of this sort do indeed provide a means of representing intended stories, then logical abduction provides a computational account of how these graphs could be constructed in the minds of people.

Applied to the Heider-Simmel film, the most-probable interpretation identified by Incremental Etcetera Abduction does indeed approximate those reported by most subjects in Heider and Simmel's early experiments. Of course, this is only the case because the commonsense axioms provided as input were carefully engineered to correctly understand the observable events in this one film. That is, the same knowledgebase would not produce human-like interpretations of any other film or narrative text. From an engineering perspective, there is not much utility to be exploited in this one application. From a scientific perspective, however, the model affords us many opportunities to explore interesting research questions. What additional information (observed or unobserved) would radically change the predominant interpretation that people have of the Heider-Simmel film? What reordering of storyline events would change the interpretation the most? By what means could the process be altered to favor interpretations that are particularly creative or clever or unusual? How might differences in commonsense knowledge of interpersonal relations determine how this film is understood across different cultures? With a computational model of narrative interpretation at hand, we are better equipped to craft testable hypotheses for each of these questions to be investigated in well-designed experiments.

4. Narrative Ambiguity

The short film crafted by Fritz Heider and Marianne Simmel is notable in that the crux of the storyline is readily inferred despite an enormous amount of ambiguity. Are the circle and the smaller triangle siblings or a romantic couple or cellmates? Is the larger triangle a wicked stepmother or a jilted ex-boyfriend or a prison guard? Is the box somebody's house, a classroom, or a prison cell? Nobody knows, and it is largely irrelevant to this story of two collaborators working to thwart their antagonist. This is the core of the story that Heider and Simmel surely aimed to tell when planning their animation, and they provided more than enough evidence in the movements of these shapes to ensure that there was minimal ambiguity as to the intended interpretation.

This notion of storytelling as ambiguity minimization can be explored further when we consider extremely ambiguous sets of observables and the possible stories that might explain them. In previous work with Ulrike Spierling [9], I investigated the application of logical abduction in story creation games, specifically the Tell Tale card game published by Blue Orange. In this casual party game, players are dealt some small number of cards depicting unrelated story elements (a monster, a deserted island, a bowl of cereal, etc.) and tasked with telling some entertaining story that weaves all the elements together into a convincing narrative. In our research, we explored how logical abduction could serve as a computational model for this type of story generation, where the cards served as the input observations to the abductive reasoning process, and the output proof graphs represented coherent narratives that explained the cards. Specifically, we considered only three cards in the Tell Tale deck, depicting a baseball player, the symbol of a heart, and a railroad train. From our friends, family, and colleagues, we obtained eight interesting stories that incorporated these elements, e.g., of a professional baseball player who falls in love with the conductor of the train he takes to practice each day, and of a baseball player at bat whose heart was racing like a train in anticipation of the pitch. For each of these eight stories, we crafted the logical axioms necessary to derive these narrative interpretations from the three cards, e.g., commonsense knowledge that being at bat in a baseball game might instill anxiety in an athlete, and that anxiety might be the cause of a racing heart.

This investigation reinforced the utility of logical abduction as a computational model of narrative interpretation, illustrating how players of the Tell Tale card game might see diverse and creative storylines in the face of extremely ambiguous observations. It also provides some insight into the larger problem of narrative generation. It is clear from this study that if some storyteller intended to unambiguously narrate any one of the eight stories we collected, that these three cards would be woefully insufficient. More narrative content would be required to steer the audience toward the intended narrative interpretation. If the story is about the anxiety of a batter in a baseball game, then the narrative might include the stern stare of the pitcher on the mound, a drop of sweat on the batter's forehead, and the thump-thump sound of a heartbeat – all elements that reinforce the intended interpretation over all other possible interpretations. In this perspective, narrative generation amounts to designing an interpretation for your audience, and then selecting the narrative content that unambiguously leads them to this destination.

5. Narratives Are Puzzles

Unlike the inkblots of Rorschach's projective psychological test, Heider and Simmel's film rewards its audience with an honest narrative, cleverly told by its creators. Despite its inherent surface ambiguities, there is enough evidence in the trajectories of its moving shapes to glean the authors' discourse intention. This process of disambiguation is fun in much the same way that Sudoku and crossword puzzles are fun, i.e., we know that there is a correct solution that we are supposed to find, and the authors have given us just enough information to guide our thinking toward this solution.

This point is illustrated by a famous six-word example of flash fiction, *"For sale: baby shoes, never worn,"* often misattributed to Ernest Hemingway as his submission in a wager among him and other writers, but more likely an adaptation of much earlier newspaper classified ads. This six-word story is meant to suggest a tragedy involving the loss of a newborn or unborn child and the extreme poverty that might compel the surviving parents to sell off the child's unused pair of shoes in a classified ad.

As a submission in a six-word story writing contest, this one is pretty good. However, its quality is entirely conditioned on the willingness of its readers to interpret it as a narrative, one that even has some vague association with a literary giant like Hemingway. If instead we treat it as random text found in a newspaper, the puzzle of the narrative disappears, leaving us with several interpretations that are much more probable. Of course there are ads for baby shoes in newspapers; probably because someone wants to make some money by selling stuff they do not need. Of course the shoes are never worn; probably because they were too small by the time a child learned to walk, or maybe the parents preferred the color of a different pair that they owned. Or maybe they are being sold as part of a clearance sale at an infant clothing store, one that does not stock used clothing. As a classified ad, these are all much more likely interpretations for these six words. It is only when we are told that it is a narrative that we treat it as a puzzle; What is the hidden interpretation that we are supposed to glean from the hints provided by the narrator? From the perspective of abductive reasoning, its identity as a narrative becomes an additional observation, a seventh word, that cannot be explained along with any of the more obvious interpretations of the other six words that are given. To be narrative, there must be a puzzle to it all, one that involves the rich emotional lives of characters pursuing their goals to comedic or tragic ends, one that was carefully designed by an author who thoughtfully selected exactly these six words to guide our thinking toward the intended interpretation.

6. Conclusions

The thesis presented in this writeup is that logical abduction is a useful computational model of narrative that supports our investigations of both narrative interpretation and narrative generation. The argument is that part of our enjoyment of narratives stems from their puzzle-like quality, where there is an intended hidden meaning for the audience to uncover that explains the words, images, or film scenes that have been presented as narrative discourse. Uncovering the intended story requires some mental labor, which we can view as a combinatorial search in the space of assumptions about the unobserved, unobservable, and unnarrated storyline events and states. We can instantiate this process in computer software that implements logical abduction, e.g., Incremental Etcetera Abduction, as was used in our explorations of the Heider-Simmel film and the Tell Tale card game.

An open research question is how best to incorporate logical abduction in computational models of narrative generation, where the starting point is the story that wants to be shared (the hidden meaning) and the aim is to model how writers, editors, filmmakers and other storytellers select the words, images, or film scenes that offer their audiences the most entertaining puzzle to solve. My intuition is that some headway can be made by considering the task of a film editor as they ponder the question: Can I safely cut this scene or line of dialogue and still have a film that provides an unambiguous path for the audience to arrive at the intended narrative interpretation? Beyond exhaustive trial and error, this hypothetical editor must have some other mental tools available to envision the audience's own abductive reasoning processes to surmise where unintended ambiguities would be created by an unfortunate cut. Perhaps this involves recognizing in the proof graph of their own interpretation of the unedited narrative the assumptions that have an overabundance of observable evidence.

Acknowledgements

Research was sponsored by the Army Research Office and was accomplished under Cooperative Agreement Number W911NF-20-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] J. Hobbs, M. Stickel, D. Appelt, P. Martin, Interpretation as Abduction, Artificial Intelligence 63:69–142, 1993.

[2] A. Gordon, Commonsense Interpretation of Triangle Behavior. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), 2016.

[3] A. Gordon, Interpretation of the Heider-Simmel Film Using Incremental Etcetera Abduction. Advances in Cognitive Systems 7, pp. 23-38, 2018.

[4] F. Heider, M. Simmel, An Experimental Study of Apparent Behavior. The American Journal of Psychology 57(2):243–259, 1944.

[5] F. Heider, The Psychology of Interpersonal Relations. Lawrence Erlbaum Associates, 1958.

[6] D. Elson, Modeling Narrative Discourse. Ph.D. Dissertation, Columbia University, 2012.

[7] T. Trabasso, P. Van Den Broek, Causal Thinking and the Representation of Narrrative Events. Journal of Memory and Language 24(5), pp. 612–630, 1985.

[8] S. Lukin, M. Walker, Narrative Variations in a Virtual Storyteller. Proceedings of the International Conference on Intelligent Virtual Agents, pp. 320–331, 2015.X

[9] A. Gordon, U. Spierling, Playing Story Creation Games With Logical Abduction. Proceedings of the Eleventh International Conference for Interactive Digital Storytelling, 2018.

A. Online Resources

Tutorial material and a Python implementation of Incremental Etcetera Abduction can be found online at https://github.com/asgordon/EtcAbductionPy