Theme Construction from Belief Conflict and Resolution

John F. Reeves
Artificial Intelligence Laboratory
3531 Boelter Hall, Computer Science Department
University of California, Los Angeles
Los Angeles, California 90024
reeves@cs.ucla.edu

Abstract

Story themes are generalized advice that a story contains, and theme recognition provides a way for a system to show that it has understood the story. THUNDER is a story understanding system that implements a model of theme construction from belief conflicts and resolutions. A belief conflict is conflicting evaluative beliefs regarding a story character's plan. When execution of the plan results in a realized success or failure for the character, a resolution to the conflict is recognized from the additional reasons that the realization provides for the evaluative beliefs in conflict. The theme of the story is generated by reasoning about how the resolution shows the beliefs in conflict to be correct or incorrect, and produces a statement of generalized advice about reasons for evaluation. Two types of advice are generated by THUNDER: (1) reason advice about the reasons for evaluation that the story shows to be correct, and (2) avoidance advice about how failures that occur as the result of erroneous evaluations could be avoided. The algorithms for constructing both type of advice and examples of THUNDER constructing themes are presented.

Introduction

Two general problems for story understanding systems are (1) knowing when a story has been understood, and (2) showing that the story has been understood. One solution is to recognize the 'theme' of the story; a story is understood when the theme is recognized, and the system shows its understanding by answering questions about what was learned from the story. The problems then become how story themes are defined, represented, constructed, and used. Additionally, a theory of theme representation and recognition identifies the components (both processes and knowledge structures) that are used in complex understanding. A computational implementation of the theory provides a independent and testable formulation of the theory and an experimental tool for creating and extending the theory.

THUNDER (THematic UNDerstanding From Ethical Reasoning) (Reeves, 1988, 1991) is a computer program that reads short narratives and answers ethical and thematic questions. For THUNDER, themes are the *generalized advice* that are constructed from conflicts and resolutions in the story. THUNDER recognizes conflicts in evaluative judgments between the reader and story characters. The resolution of a belief conflict is an event in the story that provides additional reasons for one of the beliefs in conflict, and 'shows' the 'correctness' of the belief.

To construct themes from conflicts and resolutions, THUNDER contrasts and generalizes the reasons that led to the conflict to the reasons provided by the resolution. THUNDER constructs two types of themes: (1) *reason advice* about why evaluative beliefs are correct or incorrect, and (2) *avoidance advice* about how planning failures can be avoided.

One of the stories that THUNDER reads is:

Hunting Trip

> Two men on a hunting trip captured a live rabbit. They decided to have some fun by tying a stick of dynamite to the rabbit. They lit the fuse and let it go. The rabbit ran for cover under their truck.

From *Hunting Trip*, THUNDER recognizes the following themes:

THE THEME IS THAT YOU SHOULD NOT PLAY

WITH DYNAMITE BECAUSE YOU WOULD NOT LIKE BAD THINGS TO HAPPEN TO YOU.

THE THEME IS THAT YOU SHOULD NOT EXECUTE PLANS THAT CAUSE BAD THINGS TO HAPPEN TO OTHERS FOR YOUR ENTERTAINMENT BECAUSE YOUR ENTERTAINMENT IS LESS IMPORTANT THAN BAD THINGS HAPPENING TO YOU.

THE THEME IS THAT YOU SHOULD NOT EXECUTE PLANS THAT CAUSE BAD THINGS TO HAPPEN TO OTHERS BECAUSE YOU WOULD NOT LIKE BAD THINGS TO HAPPEN TO YOU.[1]

All three of the themes that are recognized in *Hunting Trip* contain reason advice. The first theme is based on a pragmatic expectation and resolution. One of the reasons that THUNDER believed that the hunters were wrong to blow up the rabbit was because they could get hurt playing with dynamite. When the hunters' truck blows up because they were playing with dynamite, the advice in the theme is generalized from "hurting yourself" and "damaging your possessions" to "bad things happening." The second theme is based on THUNDER's reasoning that the hunters' plan was wrong because they believed that their entertainment was more important than the rabbit's life. The third theme is based on THUNDER's reasoning that the hunters' plan was wrong was because they were going to kill the rabbit. When the hunters suffer at the end of the story, they do not like it just as THUNDER did not like them blowing up the rabbit.

An example of avoidance advice is generated by THUNDER from the following *Twilight Zone* story(Day, 1985):

Four O'Clock

> Political fanatic Oliver Crangle is convinced that people who do not agree with his political views are evil. He keeps detailed files on people, makes threatening phone calls, and sends letters discrediting his 'evil' political enemies. One day, he finds a book of black magic and casts a spell to shrink every evil person in the world to a height of two feet tall at exactly four o'clock. But when the time rolls around, it is *he* who becomes two feet tall!

[1] The themes are generated in English by THUNDER when they are constructed. All I/O is verbatim from the program.

THE THEME IS THAT YOU SHOULD JUDGE YOURSELF BEFORE JUDGING OTHERS BECAUSE YOU WOULD NOT LIKE TO BE PUNISHED.

The theme is generated from the belief conflict over Oliver's plan to punish his political enemies, and the resolution of Oliver's becoming two feet tall. Avoidance advice is constructed by reasoning about how the planner could have avoided the failure. In *Four O'Clock*, Oliver was shrunk because he was guilty of the crime he was punishing others for, and the reason that he was guilty was because he was punishing others unjustly. If Oliver had evaluated his own plan, he would have avoided the failure.

THUNDER Overview

During story understanding, THUNDER's primary task is to create *evaluative beliefs* about story characters' plans. An evaluative belief is a belief that is evaluated in terms of "goodness", in contrast to a factual belief which is evaluated in terms of truth. Positive and negative evaluations of plans (called *obligation beliefs*) correspond to beliefs that the plan should or should not be used, respectively. There are two type of reasons for an evaluative belief: (1) *pragmatic* reasons, reasons about the consequences of the plan for the planner, and (2) *ethical* reasons, reasons about the consequences of the plan for others. These reasons can be broken down into two components: (1) a set of factual beliefs about the plan, and (2) a pragmatic or ethical *judgment warrant* that is used to derive an evaluative belief from factual beliefs. To generate appropriate factual beliefs about a character's plan, THUNDER has to reason about (1) plan availability, or what other options were available to the planner (2) the goal importance of successes and failures caused by the plan, both for the planner and others, and (3) the intention of the planner, and if the planner realizes that he is causing goal failures for himself or others.

THUNDER's evaluative knowledge is organized in an ideology. The representation for ideology has two components: (1) a value system, to represent what is believed to be 'good', and (2) planning strategies, to represent good ways for the goals to be achieved (Carbonell, 1980). The value system represents the goals that THUNDER believes that characters *should* try to achieve, and try not to violate. THUNDER infers the evaluative beliefs of story characters from their actions. For example, in *Hunting Trip* THUNDER

infers that the hunters believe that their entertainment is more important than the life of the rabbit. THUNDER's inference rules for evaluative belief are given in (Reeves, 1989).

From THUNDER's inferences about the beliefs of characters, THUNDER recognizes belief conflict patterns (BCPs). There are three types of belief conflicts: (1) *plan execution BCPs*, where the evaluator makes an ethical judgment that a character's plan is wrong, (2) *evaluation BCPs*, where the evaluator makes an ethical judgment that a character's reward or punishment is undeserved, and (3) *expectation BCPs*, where a character violates the evaluator's evaluative expectations. The BCP that is recognized in *Hunting Trip* is the plan execution BCP BCP:Inhumane. BCP:Inhumane represents the situation where an actor's plan is judged to be ethically wrong and (1) the plan causes non-recoverable health goal failures for another, (2) the goal failure is an integral part of the actor's plan, and (3) the goal failure is more important than the actor's goal success. The evaluation BCP BCP:No-Crime is recognized in *Four O'Clock*. BCP:No-crime represents the conflict between THUNDER's and Oliver's evaluation of Oliver's plan: Oliver believes that his plan is ethically right because punishing all evil people will protect society, while THUNDER believes that the plan is ethically wrong because he is punishing people for something that they should not be punished for.

BCPs are recognized when the plan is evaluated. Resolutions to belief conflicts are recognized when the plan is executed, and unexpected consequences (such as the hunters' truck blowing up, or Oliver shrinking) occur. The resolution provides an additional reason for the plan's evaluation. There are two types of resolutions: (1) *positive resolutions*, which are goal failures that provide additional reasons that the evaluator's belief was correct, and (2) *negative resolutions*, which are goal success that show that the planners belief was correct. The hunters' truck blowing up in *Hunting Trip*, and Oliver's shrinkage in *Four O'Clock* are instances of positive resolutions. Stories where the hunters blew up the rabbit, had a good laugh and went home, or Oliver succeeded in shirking his political opponents would have negative resolutions.

For a detailed discussion of THUNDER's natural language parser and generator, knowledge representation, inference rules, judgment warrants, and the complete set of BCPs, see (Reeves, 1991).

Theme Construction

The theme of a story is the controlling idea, central insight, unifying generalization, and purpose of the story (Perrine, 1974). To recognize the theme of the story, the reader has to identify the advice that is contained in the story, or what the story is designed to teach. The advice in the story can be an insight about life, how the world works, how to get along with others, or the reasons for or against certain courses of action.

THUNDER represents themes as (1) an abstract plan failure situation, (2) the reason for the situation's evaluation, and (3) the mistake state that led to the failure. THUNDER's representation of themes is designed to capture the following characteristics: (1) *advice* for the reader about how to plan or reason about plans so that the reader's performance will be improved, (2) *generality* so that the advice can be applied to situations that have an abstract similarity to the situation in the story, and (3) *content* that specifies the situation where the advice is to be applied and the reasons for applying the advice in the situation.

Themes in THUNDER are classified by (1) the two types of advice that the theme provides: reason advice and avoidance advice, and (2) the two types of reasons that are used to construct the theme: pragmatic and ethical. *Pragmatic themes* are advice about how to avoid planning mistakes that result in failures for the planner, and *ethical themes* are advice about why plans should not be executed because of the consequences for others.

Reason Theme Construction

A reason theme is a lesson about why a plan should not be executed. Reason themes are constructed by matching and generalizing two reasons for plan evaluation: (1) the evaluator's (THUNDER's) reason from *before* the plan is executed, and (2) the actor's reason *after* the plan is executed. The BCP recognized in the story contains the evaluator's belief and reason, and the resolution contains the actor's belief and reason.

THUNDER's reason theme construction process can be illustrated by considering what happens when the hunters' truck blows up in *Hunting Trip*. In constructing a theme, THUNDER is trying to answer the questions: (1) what does the hunters' goal failure say about why it is wrong to blow up rabbits for entertainment? and (2) how is the truck blowing up a confirmation of THUNDER's belief that blowing up the rabbit was

wrong?

When the truck blows up, the hunters evaluation of their plan changes from positive to negative because the (unexpected) loss of their truck is more important than their entertainment. The hunters' reason structure is similar to THUNDER's belief that the plan was wrong, as shown in figure 1.[2]

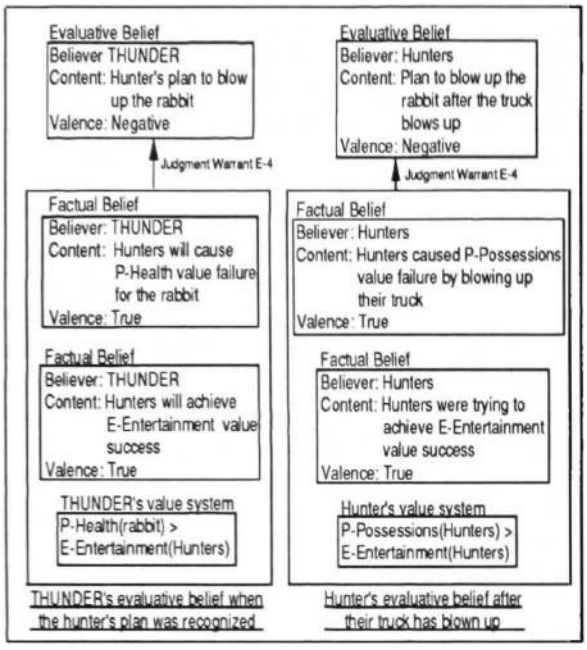

Figure 1: Beliefs used to construct the theme of *Hunting Trip*

The left side of the figure is THUNDER's belief that the hunters' are wrong to blow up the rabbit, supported by the instantiated schema for judgment warrant E-4 (If plan P1 achieves goal G while intentionally causing goal failure GF and the object of G is less important than the object of GF, then P1 is negatively evaluated). THUNDER's belief was generated when the plan was recognized after reading the second sentence of the story, and was used to recognize BCP:Inhumane. The right side of the figure shows the hunters' negative evaluation of their plan after the truck has blown up, which is also supported by judgment warrant E-4.

The matching and generalization process identifies differences in the variable bindings of the two beliefs, and generalizes the differences until the structures unify. Some variables have assigned generalizations (i.e. the believer is generalized to "you" and plan schemas are generalized to "plans"), while other types of generalizations are found by searching the item's is-a hierarchy for a common parent. If the instantiations are equal, then the instantiation is returned. For example, when the two beliefs in figure 1 are matched, the believer slot is generalized from "THUNDER" and "Hunters" to "you", and the value failure type in the top factual belief is generalized from "P-Health" and "P-Possessions" to "bad things."

THUNDER constructs the pragmatic theme from *Hunting Trip* in a similar manner. When the hunters' plan was evaluated, the TAU (Dyer, 1983) TAU:Dangerous-object was recognized that represents the potential planning failure associated with playing with dynamite ("If you play with fire, you're going to get burned"). When the reason that it is wrong to play with dynamite is matched against that hunters' pragmatic reason that the plan was wrong because their truck was destroyed, the following theme is generated:

> THE THEME IS THAT YOU SHOULD NOT PLAY WITH DYNAMITE BECAUSE YOU WOULD NOT LIKE BAD THINGS TO HAPPEN TO YOU.

The content of the belief ("playing with dynamite") is common to both THUNDER's and the hunters' beliefs. The reason in the theme ("bad things happening") is produced by generalizing from the THUNDER's expected health goal failure in TAU:Dangerous-object and the hunters' realized possessions goal failure when the truck was destroyed.

Avoidance Theme Construction

Avoidance themes are constructed by THUNDER from a *planning failure* on the part of the planner, and the planner's evaluative belief that led to the planning failure. A planning failure is a schema that is instantiated from realized goal failures, containing the action that caused the plan to fail, the action's intended effect, and the action's realized effect.

The algorithm for constructing avoidance themes is:

1. Given a BCP, resolution, and a planning failure, identify a *mistaken belief*. The mistaken belief is the belief that was used in their orig-

[2]The notation for goals is based on Schank and Abelson's goal primitives [1977]. In the notation, the goal type is signified by the letter preceding the goal name. Preservation goals (P) are motivations to keep valued acquisitions or social positions, and enjoyment (E) goals are motivation from the happiness that comes from a plan.

inal evaluation of the plan that is shown to be incorrect by the planning failure.

2. From the mistaken belief, find the part of the plan where the mistaken belief should have been recognized, and generate a new plan where the mistaken belief is checked for.

3. From the new plan, identify the failure that would have been avoided.

4. Generalize the new plan by matching the new plan to the plan executed, and generalize the reason for executing the new plan from the failure that occurred and the failure that would have been avoided.

5. Construct the avoidance theme from the generalized plan and the failure that would have been avoided.

The avoidance theme construction process is used in *Four O'Clock*. The input is (1) the BCP BCP:No-crime, (2) the resolution schema GF:Injury representing the health damage from being shrunk, and (3) Oliver's plan failure of casting the spell and shrinking himself. Since BCP:No-crime is an evaluation BCP, and THUNDER believed that Oliver's evaluation was in error, Oliver's belief that his political enemies are evil is identified as the mistaken belief. From Oliver's goal failure in the resolution, THUNDER knows that Oliver's planning error was that he failed to evaluate himself before executing the plan. By putting the step of evaluating the planner before evaluating others and generalizing the resulting structure, THUNDER generates the theme:

THE THEME IS THAT YOU SHOULD JUDGE YOURSELF BEFORE JUDGING OTHERS BECAUSE YOU WOULD NOT LIKE TO BE PUNISHED.

The avoidance theme is recognized by backtracking to find planning errors based on evaluative beliefs, and specifying the judgments that need to be made to avoid the error.

Related Work

THUNDER's model of story understanding is based on the *explanation-based* story understanding systems BORIS (Dyer, 1983) and PAM (Wilensky, 1983a). BORIS and PAM both implemented thematic knowledge structures (TAUs in BORIS, Story Points (Wilensky, 1982; Wilensky, 1983b) in PAM) to explain the events in a story in terms of what the reader learns from the story. Story points also represented the distinction between problem and solution components in stories, which is generalized in THUNDER to conflict and resolution components. THUNDER extends BORIS and PAM by (1) constructing thematic structures instead of instantiating existing schema, and (2) using evaluative belief and belief relationships as the building blocks of theme, instead of relying exclusively on planning knowledge. The representation for belief and belief relationships for ethical knowledge in THUNDER is based on the types of belief used in the OpEd system (Alvarado, 1990) for beliefs about economic plans.

The most closely related current research to THUNDER is the AQUA program (Ram, 1989). AQUA is a story understanding system that models comprehension as a goal-directed task by generating and answering questions. AQUA uses Schank's (1986) anomaly → explanation-question → explanation-pattern model where problems in understanding motivate the application of general explanations. BCPs are similar to explanation questions in that they index planning advice and potential resolutions, and motivate and direct processing. BCPs are a special class of anomaly involving evaluation that motivates explanation. AQUA takes an open-ended view of story understanding, generating many explanation questions and explanation patterns and uses the explanation patterns to learn from the story. In contrast, THUNDER takes a stratified view of story understanding where ethical judgments are used to control the understanding process. Instead of generating many explanations, THUNDER models the recognition of specific types of story themes, and uses the themes to explain the purpose of story. THUNDER and AQUA model opposing ends of the story understanding spectrum: AQUA models the inquisitive reader that ponders a story for all of its nuances, while THUNDER models the reader that recognizes one type of 'point' of the story.

Conclusions

THUNDER is a story understanding system that constructs themes from stories, based on (1) the difference between ethical and pragmatic reasons for belief conflicts and how resolutions in the story

show those reasons to be correct or incorrect, and (2) the different types of advice that can be constructed from the conflict and resolution. Ethical reasons for the belief conflict are used to generate ethical themes about how the resolution shows the plan to be right or wrong because of the consequences for others, while pragmatic reasons are used to construct pragmatic themes about the plan's consequences for the planner. THUNDER implements procedures to construct two types of advice: (1) reason advice about the reasons for evaluation that the story shows to be correct, and (2) avoidance advice about how failures that occur as the result of erroneous evaluations could be avoided.

Recognition of a belief conflict in a story is a judgment by THUNDER that something is wrong in the story. Finding a positive resolution to the belief conflict supports the evaluative belief that led to the original belief conflict. Because the story provides support for the reader's evaluation, the resolution to the belief conflict is thematic.

Recognition of a story theme provides a way for THUNDER (1) to show that it has understood the story, and (2) to know that the story has been completely processed. Identification of a theme is identification of the advice that the story was written to teach. *Thematic learning* is a two step process: (1) the theme is identified, and (2) the theme is incorporated into memory and used to improve future planning and reasoning. THUNDER accomplishes the first task, but not the second. The process of theme incorporation in THUNDER would involve adding stories and themes to episodic memory indexed by the BCP that was used to identify the theme. If the theme provides new advice, the theme can be associated with the BCP for use in future understanding. However, the theme construction algorithms provide the knowledge that will be learned.

References

Alvarado, S. J. 1990. *Understanding Editorial Text: A Computer Model of Argument Comprehension.* Kluwer Academic, Norwell, MA.

Carbonell, Jr., J. G. 1980. Towards a process model of human personality traits. *Artificial Intelligence*, 15:49–74.

Day, P. 1985. Four o'clock. In Greenberg, M. H., Matheson, R., and Waugh, C. G., editors, *The Twilight Zone: The Original Stories.* Avon Books, New York. Originally published in 1958.

Dyer, M. G. 1983. *In-Depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension.* MIT Press, Cambridge, MA.

Perrine, L. 1974. *Literature: Structure, Sound, and Sense.* Harcourt Brace Jovanovich, New York, second edition.

Ram, A. 1989. *Question-driven Understanding: An Integrated Theory of Story Understanding, Memory, and Learning.* PhD thesis, Department of Computer Science, Yale University, New Haven CT. Technical report YALEU/CSD/RR 710.

Reeves, J. F. 1988. Ethical understanding: Recognizing and using belief conflict in narrative understanding. In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-88)*, St Paul, MN.

Reeves, J. F. 1989. Computing value judgements during story understanding. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society (CogSci-89)*, Ann Arbor, MI.

Reeves, J. F. 1991. *Computational Morality: A Process Model of Belief Conflict and Resolution for Story Understanding.* PhD thesis, UCLA Artificial Intelligence Laboratory, University of California, Los Angeles.

Schank, R. C. 1986. *Explanation Patterns.* Lawrence Erlbaum, Hillsdale, NJ.

Schank, R. C. and Abelson, R. P. 1977. *Scripts, Plans, Goals, and Understanding.* Lawrence Erlbaum, Hillsdale, NJ.

Wilensky, R. 1982. Points: A theory of the structure of stories in memory. In Lehnert, W. G. and Ringle, M. H., editors, *Strategies for Natural Language Processing*, pages 345–374. Lawrence Erlbaum, Hillsdale, NJ.

Wilensky, R. 1983a. *Planning and Understanding: A Computational Approach to Human Reasoning.* Addison-Wesley, Reading, MA.

Wilensky, R. 1983b. Story grammars versus story points. *Behavioral and Brain Sciences*, 6:579–623.