Hypothesis Generation and Evidence Assembly for Intelligence Analysis: Cycorp's Noöscape Application

Nick Siegel, Blake Shepard, John Cabral, Michael Witbrock
Cycorp, Inc.
3721 Executive Center Drive, Suite 100
Austin, Texas, 78731-1615, USA
{nsiegel, blake, jcabral, witbrock}@cyc.com

Keywords: Multiple Competing Hypotheses, Question Answering, Structured Argumentation, terrorism

Abstract

This paper describes the hypothesis generation and evidence assembly capabilities of Cycorp's Noöscape application, which exploits the strengths of the Cyc knowledge base and inference engine. Noöscape allows analysts to compose questions using simple English templates, and tries to find answers via deductive reasoning. If deduction is fruitless Noöscape resorts to abduction, filling in missing pieces of logical arguments with plausible conjectures to obtain answers that are only partly supported by the facts available to the inference engine. For each conjecture, Noöscape automatically generates an augmented information retrieval (IR) query for dispatch to the analyst's preferred search engines. If the IR query results allow the analyst to determine the truth or falsity of the conjecture, a single mouse button click communicates this knowledge to Noöscape and causes the answer (and its supporting argument) to be confirmed or rejected. Noöscape presents its answers and chains of reasoning in English and provides source references for every assertion, making these key textual elements readily available for inclusion in the analyst's reports.

1. Noöscape in Use

Let us suppose that Jane Doe, an analyst, receives the following task assignment:

> Recent bombings of railways in Spain were first attributed to ETA and later to Al Qaida. Determine the most likely perpetrators of the bombings. Provide information about the motives of terrorist groups who could have carried out the bombings.

Noöscape incorporates several capabilities that will help Jane with her investigation. Upon notification that Jane has received a new assignment, for example, Noöscape automatically scans the task description and extracts terms that it recognizes, such as "ETA", "Al Qaida", "railway", "Spain", "motive", "perpetrator", "bombing", and "terrorist group". These terms are presented to Jane, who may, after review and editing, direct Noöscape to formulate augmented IR search queries for dispatch to her preferred search engines.

Noöscape also allows Jane to pose queries that take advantage of Cyc's inference engine and the knowledge sources accessible to it, including the Cyc KB. Let us assume that Jane submits this query:

[q1] What values of AGENT are there such that AGENT is a likely suspect for the March 2004 railway bombings in Spain?[1]

The inference engine produces two answers: "Basque Fatherland and Liberty" (a.k.a. "ETA") and "Al Qaida".

2. Hypothesis Generation via Abduction

When Jane poses her query, the inference engine tries to find answers by using logical deduction, employing the facts and rules contained in the KB and other accessible sources. The chain of reasoning that yields the "Al Qaida" answer runs, in part, as follows:

First, the inference engine selects rule [r1], since its conclusion matches the query:

[r1] An agent (person or organization) is a likely suspect for an action if the agent has a motive for the action, and if the agent consistently performs actions of the same type as the action in question.

The inference engine then transforms the original query into two new subqueries:

[q2] Who had a motive for the March 11 bombings?

[1] This text is Cyc's automatically generated paraphrase of the CycL version of the query. CycL is Cyc's formal, logic-based representation language. See www.cyc.com.

[q3] Who regularly carries out actions of the same type as the March 11 bombings?

The KB already contains much information about Al Qaida, including that its *modus operandi* matches the March 11 bombings. Cyc knows [f1] that Al Qaida has conducted coordinated, multiple-bomb attacks against civilians. To try to prove that Al Qaida had a motive for the March 11 bombings, the inference engine selects this rule:

[r2] An agent has a motive for an action if the agent has a negative vested interest in the policy of some organization, and the action causes an immediate decrease in the popularity of the organization.

The KB contains these relevant facts:

[f2] Al Qaida is a terrorist organization (a type of intelligent agent);
[f3] The Spanish People's Party has a policy of including Spain in the "coalition of the willing"; and
[f4] Al Qaida opposed (*i.e.*, had a negative vested interest in) this policy.

Together, [f2], [f3], and [f4] are almost sufficient to satisfy the conditional part of [r2], but at the time of Jane's query the KB contains no information about the effect of the March 11 bombings on the popularity of the Spanish People's Party. No deductive tactic is successful at this point in the inference. Therefore, the inference engine introduces a hypothetical statement to complete the chain of reasoning:

[h1] The March 11 bombings cause an immediate decline in the popularity of the Spanish People's Party.

This tactic allows the inference engine to return "Al Qaida" as a plausible but strictly provisional answer. It is far better than no answer at all, but its validity hinges on the truth or falsity of conjecture [h1]. Thus, by introducing [h1], the inference engine identifies a new topic of inquiry for the analyst. The chain of reasoning that results in the "ETA" answer also relies on a conjecture that identifies a new topic for research:

[h2] ETA regularly carries out coordinated, multiple-bomb attacks on civilians.

To generate opportune hypothetical statements such as those described above, Cyc's inference engine uses a form of reasoning called *abduction* (Paul 1993). In simplest terms, abductive inference follows this pattern: If **q** is known to be true and **p => q** is also known to be true, then posit **p** as a hypothetical support, or explanation, for **q**. This mode of reasoning transforms every rule in the KB (at least potentially) into a template for hypothesis generation. The inference engine's ability to combine deduction and abduction in a general purpose, logic-based question answering procedure is readily applicable to some of the evidence assembly and argument construction tasks faced by analysts. In Noöscape, this ability is always turned on.[2]

3. Justifications and Hypothesis Assessment

Noöscape displays English versions of the proofs, or chains of reasoning, that support its answers. A single answer may be supported by multiple proofs. The analyst can selectively expand the amount of information displayed, or can collapse an entire proof to a summary. For every statement in a proof, Noöscape shows the name of the person or entity who entered the statement into the KB, the date and time of entry, and citations of supporting data repositories, written sources (with active hyperlinks, when possible), or named experts (with contact information, when possible).

Noöscape indicates those statements in a proof that are conjectures, and allows the analyst to resolve the hypothetical status of these statements. If the analyst does not know whether a given conjecture is true or false, Noöscape can dispatch a web search query automatically generated from the conjecture, the results of which may allow the analyst to determine its veracity. The analyst is encouraged to cite supporting sources when resolving a conjecture. Answers, proofs, and source citations can easily be exported for inclusion in the analyst's reports.

4. Significance

Noöscape expands the scope of automated question answering by combining multi-faceted, logic-based reasoning techniques with IR search to provide benefit whenever possible. Conclusions are supported by deductive proofs, or by arguments that include conjectures and motivate new topics of inquiry. Proofs are backed by source citations, with active links to the sources themselves when possible.

Acknowledgements

Support for the development of Noöscape thus far, and for the writing of this paper, has been provided by ARDA's NIMD program.

References

Ng, H. and Mooney, R. 1990. On the Role of Coherence in Abductive Explanation. *Proceedings of the Eighth National Conference on Artificial Intelligence,* 337-342. Boston, MA.

Paul, G. 1993. Approaches to Abductive Reasoning: An Overview. *Artificial Intelligence Review* 7:109-152.

[2] See Ng and Mooney (1990) for a concise discussion of varieties of abductive inference and control mechanisms.