Modeling Commonsense Reasoning via Analogical Chaining: A Preliminary Report

Joseph A. Blass (joeblass@u.northwestern.edu)
Kenneth D. Forbus (forbus@northwestern.edu)
Northwestern University, 2133 Sheridan Road,
Evanston, IL 60208

Abstract

Understanding the nature of commonsense reasoning is one of the deepest questions of cognitive science. Prior work has proposed analogy as a mechanism for commonsense reasoning, with prior simulations focusing on reasoning about continuous behavior of physical systems. This paper examines how analogy might be used in commonsense more broadly. The two contributions are (1) the idea of *common sense units*, intermediate-sized collections of facts extracted from experience (including cultural experience) which improves analogical retrieval and simplifies inferencing, and (2) *analogical chaining*, where multiple rounds of analogical retrieval and mapping are used to rapidly construct explanations and predictions. We illustrate these ideas via an implemented computational model, tested on examples from an independently-developed test of commonsense reasoning.

Keywords: Analogical Reasoning; Commonsense Reasoning; Analogical Abduction

Introduction

Consider three situations. (1) A person throws a crumpled-up piece of paper the size of an egg at another person's head. (2) is like (1), but the item thrown is an actual egg. (3) is like (1), but the item is a small, white stone of the same size. Despite these situations' similarities, you would likely interpret the first as a playful action, the second as emotionally aggressive but perhaps not too harmful, and the third as a serious act of aggression. These conclusions come quickly and easily to us, without conscious pondering. Such extremely rapid construction of explanations and predictions is a hallmark of commonsense reasoning.

Several models for commonsense reasoning have been proposed, ranging from logical reasoning using general, first-principles axioms (e.g. Davis, 1990, Lenat, 1995) to numerical simulation (e.g. Battaglia *et al*., 2013). We take analogical reasoning as a promising approach for explaining commonsense reasoning, for three reasons. First, analogical reasoning can work with partial knowledge: we may not have a fully articulated general theory of how much harm being hit by something might have, but if we have examples, we can still work with those. Second, analogical generalization provides a potential mechanism for learning probabilistic generalizations to represent experience. Third, analogy can allow a system to generate multiple inferences by importing whole relational structures from a single case, rather than requiring separate rules for each inference.

Our prior work on exploring analogy in commonsense reasoning focused on reasoning about the behavior of continuous systems (e.g. Forbus & Gentner, 1997; Forbus 2001). Here we explore how analogy might be used for commonsense more broadly. We have argued that much of human abduction and prediction might be explained by analogy over experiences and generalizations constructed from experience (Forbus, 2015). This paper explores in more detail how that might work. Specifically, we propose that multiple analogical retrievals are used to quickly elaborate a situation, providing a set of plausible explanations and predictions. We call this process *analogical chaining* (AC). The units that are retrieved might be specific situations or larger structures, such as traditional scripts (e.g. Schank & Abelson, 1977) and frames (e.g. Minsky, 1974), if they are good matches for the situation. However, we also propose that experience is factored into *Common Sense Units*, cases in the case-based reasoning sense, that are typically larger than single facts and smaller than frames or scripts[1]. A CSU consists of several facts that connect, for example, an event of a particular type with a precursor or with a potential outcome. Such cases can be useful for prediction when the precursor matches the current situation, and for explanation when the outcome matches the current situation (Forbus, 2015). Because they are smaller, they should be more easily transferrable to a wider range of situations, because they contain less non-overlapping information.

We begin by reviewing the structure-mapping models that this model is built upon, and the Cyc-derived ontology used. We describe our model, including our hypotheses about the nature of CSUs and the computational issues raised by AC. We present an experiment where a pool of CSUs are used to answer questions from the Choice of Plausible Alternatives (COPA, Roemmele *et al*., 2011) test of commonsense reasoning. We close with related and future work.

Background

Analogy is an important tool for reasoning and decision-making; we use past experiences to understand and make decisions in new situations (Markman & Medin 2002). We use Gentner's (1983) structure-mapping theory of analogical reasoning, which argues that analogy involves finding an alignment between two structured descriptions. The Structure-Mapping Engine (SME, Forbus et al. 2016) is a computational model of analogy and similarity based on structure mapping theory[2]. SME takes in two structured, relational cases (a *base* and a *target*) and computes up to three *mappings* between them. These include the correspondence between the two cases, *candidate inferences* suggested by the

[1] We are inspired by the Goldilocks Effect argument for analogical reasoning (Finlayson & Winston, 2005).

[2] See (Gentner & Forbus, 2011) for a survey of other models.

mapping, and a *similarity score* that serves as a measure of how good the mapping is. If a candidate inference involves an entity not present in the other case, that entity is hypothesized as a *skolem entity*.

Running SME across every case in memory would be prohibitively expensive, and implausible for human-scale memories. MAC/FAC (Forbus et al. 1995) retrieves cases that may be helpful for analogical reasoning from a case library, without relying on any indexing scheme. It takes in a *probe* case like those used by SME, as well as a case library of other such cases. MAC/FAC efficiently generates remindings, which are SME mappings, for the probe case with the most similar case retrieved from the case library. MAC/FAC proceeds in two stages: first, it computes dot products between content vectors of the probe and each case in the case library, a coarse approximation for scalability. Up to the three most similar cases are passed to the second stage, which uses SME to calculate similarity based on structured descriptions. Typically only one, but up to three if close, retrievals are output by MAC/FAC.

We use the Cyc ontology (Lenat, 1995) as a source of representations. The subset of contents of ResearchCyc that we use for our knowledge base contains over 110,000 concepts and over 32,000 relations, constrained by over 3.8 million facts. We extend this knowledge base to support qualitative reasoning, analogical reasoning and learning, and additional lexical and semantic information. The knowledge is partitioned into over 58,000 *microtheories*, which can be linked via inheritance relationships to form logical environments to support and control reasoning. The MiddleEarthMt or other microtheories representing fictional worlds, for example, are rarely useful in commonsense reasoning (although the converse is not true). Microtheories simplify the consideration of alternatives during reasoning.

Using ResearchCyc representations allows us to leverage the several person-centuries of work that has gone into its development, but also reduces the risk of tailorability. By using natural language inputs and someone else's representations, we reduce the chance that our results are an effect of having spoon-fed answers to our systems.

Common Sense Units

People are spontaneously reminded of similar prior situations. We further hypothesize that experience, both direct and culturally transmitted (e.g., reading, watching videos) is carved up into smaller pieces as well, and combined via analogical generalization to create probabilistic structures (via SAGE, McLure et al. 2015). These lack logical variables but can behave like rules when applied by analogy, and serve as grist for analogical reasoning about novel situations. Because they include fewer statements they are less specific (in the model theory sense), and more likely to match to a wide range of cases.

We think of CSUs as the set of facts surrounding a particular common plausible inference. For example, a CSU for reciprocity might encode simply that one agent performs a positive deed for another, which causes the other agent to perform a positive deed for the first at some future time. CSUs are intended to be smaller than situations, hence making them more compositional. We have not yet explored learning CSUs, because we first want to establish that they can be useful. The current paper provides evidence for this.

Analogical Chaining (AC)

Many prior computational models of analogical reasoning have treated analogy as a one-shot process, where a single analog is retrieved and used, or perhaps replaced with another if the first is not satisfactory. We go beyond that here by chaining analogies, i.e. using the elaboration of a situation via analogical inferences to retrieve yet more analogs, similar to how chaining in logical inference works. This is conceptually similar to Derivational Analogy (Veloso & Carbonell, 1993; differences discussed below).

AC proceeds as follows. The case library of CSUs is a stand-in for some of the commonsense knowledge a human gains over their lifetime. The current situation (the *target*) is used as a probe for MAC/FAC over the case library. If no mapping is produced, the program seeks another reminding, without cases that were rejected or previously used. If a mapping is found, any candidate inferences are asserted into an inference context, along with statements indicating what category any skolems belong to. Inferences are placed in a separate context from the case because there is no guarantee that they are correct. Another retrieval is then performed, with the probe being the union of the target and the inference context. If no information was added to the case, the previously retrieved analog is suppressed, to prevent looping. When information is added to the inference context, previously rejected CSUs are freed up to be retrieved against in case they might build off the inferences made. The process repeats until an answer has been found (for a question-answering task) or there are no more inferences to carry over into the target case (Figure 1).

There are several potential advantages to this model. Cases can be dynamically added to the case library, and can be used immediately. AC enables both inference about what is present in the case (filling in implicit relational links) as well as abductive explanations for what caused an event or predictions about what might happen next.

The strongest advantage of analogical reasoning is that, unlike logical inference, it does not require a fully articulated logically quantified theory. The difficulty in creating such theories is well-known, and seems to stem from two reasons. First, people have difficulty articulating a complete, accurate account of their reasoning. Second, their models tend to be full of gaps and unintended consequences. By contrast, reasoning by analogy from experience does not require a complete axiomatic theory of, for example, causality or human actions. It only requires examples with explanations specific to those cases. Analogical reasoning moreover is guided by what has happened, rather than what might be logically possible. To be sure, analogy can go awry as well – no powerful reasoning system with imperfect information and finite resources can always guarantee valid results. AC

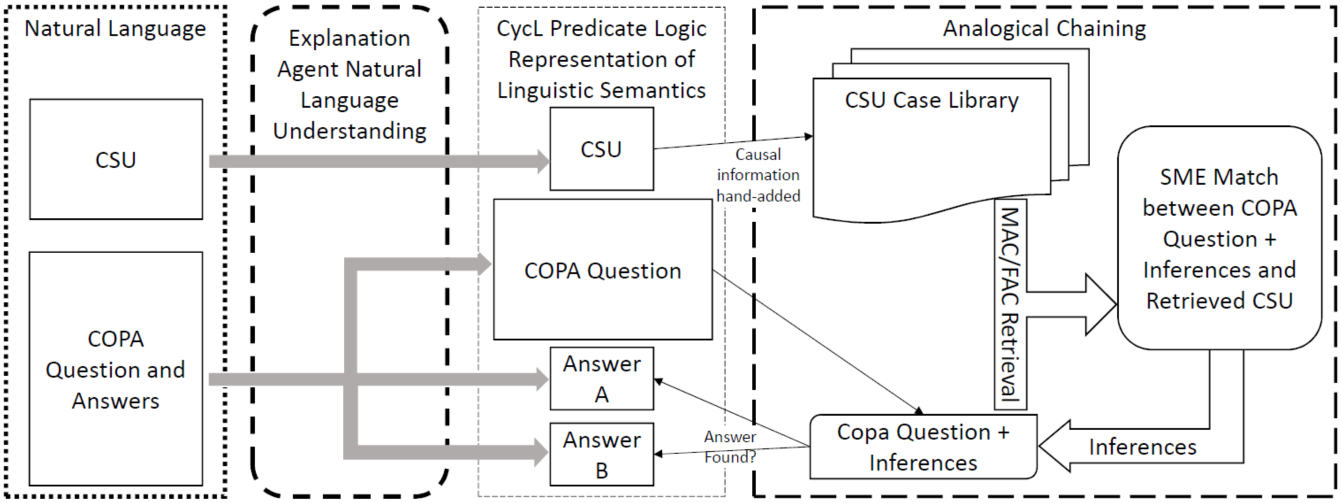

Figure 1: Analogical Chaining Workflow for answering COPA questions

should provide a compression of the inference space, both in terms of the number of inferences completed per step and fewer inappropriate branches explored, compared to logical chaining.

Simulation

Method

To explore the plausibility of these ideas, we implemented AC using the Companion cognitive architecture (Forbus *et al.* 2009). For testing, we focused on a small subset of COPA training set questions[3], and automatically encoded the questions and the majority of the CSUs via natural language understanding capabilities built into the architecture. These include six questions involving the causes and consequences of situations involving violent impacts, and a seventh question involving boiling water. These questions are shown in Figure 2. This paper uses question 461 for illustration.

```
214: The vandals threw a rock at the window.
What happened as a RESULT?
    The window [cracked / fogged up].
294: The egg splattered. What was the CAUSE of
this?
    I [dropped / boiled] it.
347: The boy got a black eye. What was the CAUSE
of this?
    The bully [mocked /punched] the boy.
370: The water in the teapot started to boil.
What happened as a RESULT?
    The teapot [cooled / whistled].
390: The truck crashed into the motorcycle on
the bridge. What happened as a RESULT?
  [The motorcyclist died / The bridge collapsed].
461: The mother called an ambulance. What was
the CAUSE of this?
 Her son [lost his cat / fell out of his bed].
496: My ears were ringing. What was the CAUSE of
this?
    I went to a [museum / concert].
```

Figure 2: The Selected COPA Questions and Answers

[3] We use training set questions here because publishing test set questions would violate the security of the test.

We created 32 CSUs designed to be relevant to the topics of the questions, plus distractors. These CSUs ranged in size from 2 to 8 facts. COPA questions are designed so that both answers are actually plausible, but one answer is always more plausible than the other. Consequently, CSUs that would contribute to incorrect answers were encoded as part of the set, as well as several CSUs irrelevant to answering the specific questions chosen (for example, that a system with a faulty component may malfunction). Sample CSUs are shown in Figure 3. Representations for 19 CSUs were almost

When a loved one is hurt, you call an ambulance.

```
(loves caller6829 person6293)
(senderOfInfo call22246 caller6829)
(communicationTarget call22246 ambulance22371)
(isa ambulance22371 Ambulance)
(isa call22246 MakingAPhoneCall)
(causes-PropProp
 (and (objectHarmed someHarm1523 person6293)
      (loves caller6829 person6293))
 (and (isa call22246 MakingAPhoneCall)
      (senderOfInfo call22246 caller6829)
      (communicationTarget call22246 ambulance22371)
      (isa ambulance22371 Ambulance)))
                    *********
```

Mothers love their sons (similar CSUs cover mothers & daughters)

```
(isa mother22349 HumanMother)
(sons mother22349 son26849)
(loves mother22349 son26849)
(causes-PropProp
 (and (isa mother22349 HumanMother)
      (sons mother22349 son26849))
 (loves mother22349 son26849))
                    *********
```

Falling out of bed hurts.

```
(isa bed2498 Bed-PieceOfFurniture)
(isa fall24789 FallingEvent)
(isa impact1953 ViolentImpact)
(objectHarmed impact1953 person22386)
(primaryObjectMoving fall24789 person22386)
(causes-PropProp
 (and (isa fall24789 FallingEvent)
     (from-UnderspecifiedLocation bed2498 person22386)
      (isa bed2498 Bed-PieceOfFurniture)
      (primaryObjectMoving fall24789 person22386))
 (and (isa impact1953 ViolentImpact)
      (objectHarmed impact1953 person22386)))
```

Figure 3: CSUs required to solve COPA question 461.

```
(communicationTarget call22246 ambulance22371)
(isa ambulance22371 Ambulance)
(isa call22246 MakingAPhoneCall)
(loves caller6829 person6293)
(senderOfInfo call22246 caller6829)
(causes-Underspecified love9172 call22246)
                    *********
(causes-PropProp
 (and (objectHarmed someHarm1523 person6293)
      (loves caller6829 person6293))
 (and (communicationTarget call22246 ambulance22371)
      (isa ambulance22371 Ambulance)
      (isa call22246 MakingAPhoneCall)
      (senderOfInfo call22246 caller6829)))
```

Figure 4: Top: the NLU-output CSU about calling an ambulance when a loved one is hurt. Bottom: the causal fact after manual editing (other facts the same)

entirely automatically generated by our NLU system, with only causal representations manually edited (described next). The remaining 13 CSUs also began as automatically processed natural language, but required more significant manual changes, due to various limitations of the NLU system, mostly involving words unknown to the system.

Causal representations automatically generated from natural language were modified by hand. The NLU system generates structurally flat causal representations, which are difficult for SME to operate over. For example, saying that someone you loved being hurt leads to calling them an ambulance results in an underspecified causal relation. That information was automatically extracted by the NLU system but not connected to the causal fact; we edited those facts to connect those relevant automatically generated facts (Figure 4). Additionally, we added facts to several CSUs indicating that a particular event was an instance of a ViolentImpact, (a new concept for our system), and removed facts which made the CSU overly specific (i.e., information that would be worn away via analogical generalization).

Since AC involves within-domain analogies, we use required partition constraints (Forbus *et al*., 2016) to restrict matching entities to be within the same categories. For example, matches with the CSU in Figure 3 had the constraint that ambulance and phone call could only be placed in correspondence with an ambulance and a phone call, respectively.

```
(causes-PropProp
 (and (isa rock2942-skolem StoneStuff)
      (isa throw2912-skolem ThrowingAnObject)
      (objectActedOn throw2912-skolem rock2942-skolem)
      (target throw2912-skolem person6293-skolem))
 (and (isa someHarm1523-skolem ViolentImpact)
      (objectHarmed someHarm1523-skolem
                    person6293-skolem)))
                    *********
(causes-PropProp
 (and (from-UnderspecifiedLocation bed2498-skolem
                                   person6293-skolem)
      (isa bed2498-skolem Bed-PieceOfFurniture)
      (isa fall24789-skolem FallingEvent)
      (primaryObjectMoving fall24789-skolem
                           person6293-skolem))
 (and (isa someHarm1523-skolem ViolentImpact)
      (objectHarmed someHarm1523-skolem
                    person6293-skolem)))
```

Figure 5: Plausible inferences for question 461: the person was hit by a rock; the person fell out of bed (correct)

For each question, the Companion read the question and answers into separate microtheories. The system read and understood the questions without human intervention. The Companion automatically filtered out the phrase asking for cause or effect, since we found that for most COPA questions only one answer is plausible regardless of which is asked for. The Companion then used AC to flesh out the question, storing the inferences in a separate microtheory.

After each extension, the Companion would check whether it had reasoned its way to one of the answers, using SME. The base normalized score (i.e. the similarity of the base/target divided by the similarity of the base with itself) measures how much of the base is covered by the match. Here, an answer is used as the base and the union of the question and inferences microtheories are used as the target. If the base normalized score of the comparison is above 0.999, all the facts in the answer have identical (but for entity tokens) corresponding facts in the reasoning microtheory, and the model selects that answer as correct.

Results

Of the seven questions selected, a Companion using AC was able to answer six correctly. Most inferences generated through chaining either helped the system find the answer or were perfectly plausible (Figure 5), although in some cases it considered at least one strange possibility before finding the right answer (Figure 6). Answering five of these six questions correctly involved chaining through the same CSU expressing that a violent impact causes harm, demonstrating that AC can use the same CSU in different contexts.

Question 461 was the only question which was not answered correctly from its raw NLU output, which included representations that prevented SME from detecting success. Specifically, in the correct answer "her son fell out of his bed," the multiple possessives "her" and "his" were interpreted as (possesses mother son) and (possesses his bed), which are reasonable. However, the coreference system did not resolve "his" to "son" (i.e., (possesses son bed)), and the CSU did not contain the first "possesses" fact (another fact in the CSU expressed the mother/son relationship), so the base normalized score of the match was not quite high enough to detect success. However, there was still the information that the son fell out of *a* bed, if not *his* bed. To verify this analysis, we removed these extra "possesses" facts, and the system was able to correctly find the answer.

Was analogical chaining necessary? Yes, since every question required two or three analogies to reach the correct answer. For example, after amending question 461 as noted

```
(causes-PropProp
    (and (isa someHarm1523-skolem Concert)
         (isa go-to35116-skolem AttendingSomething)
     (toLocation go-to35116-skolem someHarm1523-skolem)
      (performedBy go-to35116-skolem person5082-skolem)
         (loudnessOfEvent someHarm1523-skolem Loud)))
    (and (isa someHarm1523-skolem ViolentImpact)
         (objectActedOn someHarm1523-skolem
                        ear2942-skolem)
         (isa ear2942-skolem Ear)))
```

Figure 6: Implausible inference for question 461: the harm was a concert which hurt their ears

above, the system was able to find the answer only after retrieving and applying the three CSUs in Figure 2. It first postulated that a loved one had been hurt, then that it was her son, and from a fall. We suspect that most COPA questions should be answerable within three AC steps, but confirming this remains future work.

Related Work

As of this writing, three other systems have been tested on COPA, all focused on text analysis. Gordon *et al*. (2011) used Pointwise Mutual Information to evaluate how often words in the questions co-occurred with words in the answer. While their system performed significantly above chance (65.4% accuracy), it only slightly gained in accuracy as the training corpus dramatically increased, from 10^6 to 10^7 stories. Goodwin *et al*. (2012) achieved a similar performance with other textual analysis techniques (63.4% accuracy), but found that using multiple components in their analysis did not significantly improve accuracy over using only bigram co-occurrence. Luo *et al.* (2016) achieved higher accuracy (70.2%) using a large corpus to automatically extract causal relationships between concepts, then using this extracted information to determine the 'causal strength' between a question and each of its answers. While the extracted causal information appears more effective than the other two techniques, it still requires that information to be represented in the training corpus, which much of commonsense knowledge is not. Together these findings suggest that there are upper limits to text-based techniques, which argues for investigating approaches like ours that use conceptual representations. Of course, all three of these techniques were able to attempt the entire COPA test. AC will require a large case library of CSUs before we test it on the full COPA test.

Derivational analogy, as implemented in the PRODIGY architecture, similarly chains together previously known cases to derive solutions to problems (Veloso & Carbonell, 1993). It plans for a goal by analogy to plans that previously achieved a similar goal, with subgoals recursively planned for by analogy. Stored cases are indexed by and retrieved via their justifications, initials states, and goal states. Derivational analogy differs from AC as we have described it in three important ways. First, our cases are stored and retrieved without requiring any information about what they previously allowed the system to conclude. Although this means that sometimes a highly dissimilar yet nonetheless relevant case may not be retrieved by MAC/FAC, it also circumvents issues with indexing and retrieval, and enables AC to use a relevant case even when it has not been useful in similar past situations. Second, derivational analogy was specifically used to create plans to achieve goals, rather than to explain a state of affairs or predict future outcomes. It is not clear whether derivational analogy could be used for tasks that cannot be easily framed in terms of planning or problem-solving (although answering COPA questions could be framed in such a way). Finally, PRODIGY made use of both case-based and first-principles reasoning. AC does not use any first-principles reasoning at any stage.

Much AI research on commonsense reasoning has relied on formal logic and deductive inference (see Davis, 1990 and Mueller, 2014). All such approaches rely on using large numbers of logically quantified axioms. We have noted several problems with this approach, including the difficulty of constructing correct logically quantified axioms. Analogy only requires acquiring relevant cases and refining them via analogical generalization, rather than complete and correct domain theories. Furthermore, reasoning using formal logic must proceed serially: each inference rule asserts only its consequences. AC also proceeds serially, but a highly relevant case can lead to several inferences (not necessarily derivable from the same logical rule) being asserted at once, potentially reducing the number of inference steps needed.

In Explanation-Based Learning (EBL) (DeJong, 2006), a human provides a formal domain theory and examples from a domain to a system, which it uses to refine its own formal theory of that domain. AC differs in that it only requires the human to provide (in simplified natural language) cases that illustrate an underlying principle, rather than the logic of that underlying principle, which is simpler for non-experts. Also, the domain theories generated through EBL are still in logic and as such face the same drawbacks listed above.

AnalogySpace (Speer, Havasi & Lierberman 2008) used a large knowledge base of commonsense assertions in natural language to make predictions about concepts, which could then be compounded with further predictions. However, they define similarity as a linear operation over feature vectors, using a reduced-dimensionality approximation of MAC's dot-product of content vectors to retrieve relevant concepts, and do not include any measure of structural similarity. Furthermore, this work was only used to predict features of individual concepts, and it is unclear how it would scale up to explain or predict larger cases.

Though AC generates possible explanations for situations, it differs from using logic for abduction (e.g. Hobbs, 2006) since it does not require a logically quantified domain theory, and does prediction as well as explanation.

The importance of the Goldilocks Principle, using cases that are neither too small or too large in analogical matching, was highlighted by Finlayson and Winston (2005), which helped inspire our thinking about CSUs..

Conclusions and Future Work

There is already evidence that analogy is widely used in human cognition (Gentner, 2003), so it would be surprising if it were not used for commonsense reasoning. This paper has explored how that might work. We proposed Common Sense Units, intermediate-sized representations, closer to rules in size than raw experiences, but still without logical variables, as a means of encoding experience that supports flexible analogical processing. We proposed analogical chaining, the repeated use of analogies to rapidly construct explanations and predications, as a means of performing commonsense reasoning. While AC is serial at the level of applying an analog, it is parallel with respect to the application of candidate inferences within a step, thereby being more efficient than traditional chaining with logical axioms. The bundling of common patterns of facts via CSUs also provides more focus for each inference step. CSUs and AC were used to answer COPA questions, demonstrating its potential as a model of commonsense reasoning.

We plan to explore several directions of future work. First, we plan to expand our NLU capabilities to support fully automatic construction of CSUs from natural language, rather than mixing automatic generation with some manual editing, both to reduce tailorability and to expand coverage, including crowdsourcing CSUs (c.f. Li *et al*., 2013). Extracting CSUs from larger stories via analogical generalization is, we think, a promising approach. Second, we plan to expand the reasoning techniques used for checking the validity of retrieved cases, skolem resolution, and determining when sufficient reasoning has been done. Answers to multiple-choice questions can also help guide chaining. Finally, we plan to test the expanded model on the entire COPA and other commonsense reasoning tests, such as Winograd schemas[4].

Acknowledgements

This research was supported by the Socio-Cognitive Architectures for Adaptable Autonomous Systems Program of the Office of Naval Research, N00014-13-1-0470.

References

Battaglia, P., Hamrick, J., & Tenenbaum, J. (2013). Simulation as an engine of physical scene understanding. *PNAS*, *110*(45), 18327-18332.

Davis, E. (1990, 2014). *Representations of commonsense knowledge*. Morgan Kaufmann.

DeJong, G. (2006). Toward robust real-world inference: A new perspective on explanation-based learning. In *Machine Learning: ECML 2006* (pp. 102-113). Springer.

Finlayson, M. A., & Winston, P. H. (2005) Intermediate Features and Informational-level Constraint on Analogical Retrieval. In *Proceedings of the 27th Annual Meeting of the Cog. Sci. Society (CogSci 2005)*, Stresa, Italy. 666–671.

Forbus, K. 2001. Exploring analogy in the large. In Gentner, Holyoak, and Kokinov (Eds) *The Analogical Mind: Perspectives from Cog. Sci.* Cambridge, MA: MIT Press.

Forbus, K. (2015). Analogical Abduction and Prediction: Their Impact on Deception. *AAAI Fall Symposium on Deceptive and Counter-Deceptive Machines.*

Forbus, K., Ferguson, R., Lovett, A., & Gentner, D. (2016). Extending SME to handle large-scale cognitive modeling. *Cognitive Science.*

Forbus, K. & Gentner, D. 1997. Qualitative mental models: Simulations or memories? *QR 1997*, Cortona, Italy.

Forbus, K., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*(2), 141-205.

Forbus, K., Klenk, M. and Hinrichs, T. (2009). Companion Cognitive Systems: Design Goals and Lessons Learned So Far. *IEEE Intelligent Systems*, 24(4), 36-46.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Sci.*, *7*(2), 155-170.

Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp.195-235). Cambridge, MA: MIT Press.

Gentner, D., & Forbus, K. (2011). Computational models of analogy. *WIREs Cognitive Science*, 2. 266-276.

Gentner, D., Ratterman, M. J., & Forbus, K. (1993). The roles of similarity in transfer. *Cog. Psych.,* 25, 524–575.

Gordon, A., Bejan, C., and Sagae, K. (2011) Commonsense Causal Reasoning Using Millions of Personal Stories. *Twenty-Fifth Conference on Artificial Intelligence (AAAI-11),* August 7–11, 2011, San Francisco, CA

Goodwin, T., Rink, B., Roberts, K., and Harabagiu, S. (2012) UTDHLT: COPACETIC System for Choosing Plausible Alternatives. *Procs of the 6th Int'l Workshop on Semantic Eval.* (SemEval 2012), June 7-8, 2012, Montreal, Canada.

Hobbs, J. (2006) Abduction in Natural Language Understanding. In Horn & Ward (Eds.)., *The Handbook of Pragmatics* Blackwell Publishing Ltd, Oxford, UK.

Lenat, D. (1995). CYC: A large-scale investment in knowledge infrastructure. *Comm. of ACM*, *38*(11), 33-38.

Li, B., Lee-Urban, S., Johnston, G., & Riedl, M. (2013). Story Generation with Crowdsourced Plot Graphs. In *Proceeding of the 27th AAAI Conference on Artificial Intelligence*, Bellevue, Washington.

Luo, Z., Sha, Y., Zhu, K., Hwang, S., & Wang, Z. (2016). Commonsense Causal Reasoning between Short Texts. *15th Int'l Conference on Principles of Knowledge Representation and Reasoning*, Cape Town, South Africa.

Markman, A. B., & Medin, D. L. (2002). Decision making. *Stevens' Handbook of Experimental Psychology*.

McLure, M. D., Friedman, S. E., & Forbus, K. D. (2015). Extending Analogical Generalization with Near-Misses. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, Austin, TX (pp. 565-571).

Minsky, M. (1974). A Framework for Representing Knowledge. Reprinted in *The Psychology of Computer Vision,* P. Winston (Ed.), McGraw-Hill, 1975.

Mueller, E. T. (2014). *Commonsense Reasoning: An Event Calculus Based Approach*. Morgan Kaufmann.

Roemmele, M., Bejan, C., & Gordon, A. (2011) Choice of Plausible Alternatives: An Evaluation of Commonsense Causal Reasoning. *AAAI Spr. Symp. on Log. Form. of Commonsense Reasoning*, Stanford University

Schank, R.C. & Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding*. Hillsdale , NJ: Earlbaum Assoc.

Speer, R., Havasi, C., & Lierberman, H. (2008). AnalogySpace: Reducing the Dimensionality of Common Sense Knowledge. *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*, Chicago, IL

Veloso, M., & Carbonell, J. (1993). Derivational analogy in PRODIGY: Automating case acquisition, storage, and utilization. In *Case-Based Learning* (55-84). Springer US.

[4] http://commonsensereasoning.org/winograd.html