Practical Optimal-Solution Algorithms for Schema-based Analogy Mapping

Robert Hamilton (rhamilton@mun.ca) and **Todd Wareham (harold@mun.ca)**
Department of Computer Science
Memorial University of Newfoundland
St. John's, NL Canada AIB 3X5

Keywords: Analogy Mapping; Computational Complexity

Schema-based Analogy Mapping

In the last 20 years, analogy derivation has come to be at the forefront of cognitive science (Gentner, 2010). Under Structure Mapping Theory (Gentner, 1983), an **analogy** "T is (like) a B", where B and T are predicate-structures, is a mapping from a portion of B to a portion of T that satisfies various conditions (see Figure 1). This yields the following computational problem:

> ANALOGY MAPPING
> *Input:* Two predicate-structures B and T.
> *Output:* The most systematic analogy mapping M between B and T.

One popular heuristic for making this problem simpler is to use schemas. While there is no formal definition, a schema is described in Gentner et al. (2009) as "...the relational structure engendered by an analogical comparison ... [which] will be a fairly concentrated relational representation, with many of the initial item-specific features stripped away" (p. 1345). It has been claimed that schema-based analogy derivation is efficient in practice, i.e.

> ...aligning a target with [a schema] should be computationally less costly than aligning a target with the corresponding literal base concept [because schemas] will contain fewer predicates than the literal concepts they were derived from, and a higher proportion of these predicates can be mapped to relevant target concepts. (Bowdle & Gentner, 2005, p. 199)

The common thread in such claims is that schemas make analogy derivation easier because schemas are small and will generally be fully or almost-fully mapped to a comparison predicate-structure. However, these claims have never been formally proven. In this poster, we give preliminary results of a complexity-theoretic investigation of these claims.

Methodology

First, we establish the complexity of ANALOGY MAPPING. Then, we analyze a simplified version of this problem dealing specifically with schemas in order to find efficient algorithms. Following convention in Computer (Garey & Johnson, 1979) and Cognitive (van Rooij, 2008) Science, an algorithm is considered efficient if it runs in polynomial time, i.e, in time upper-bounded by n^c where n is the input size and c is a constant. It is widely held that no such algorithm exists for a problem if that problem is NP-hard. In such cases, we consider two ways of restricting the problem input to allow for a practical solution:

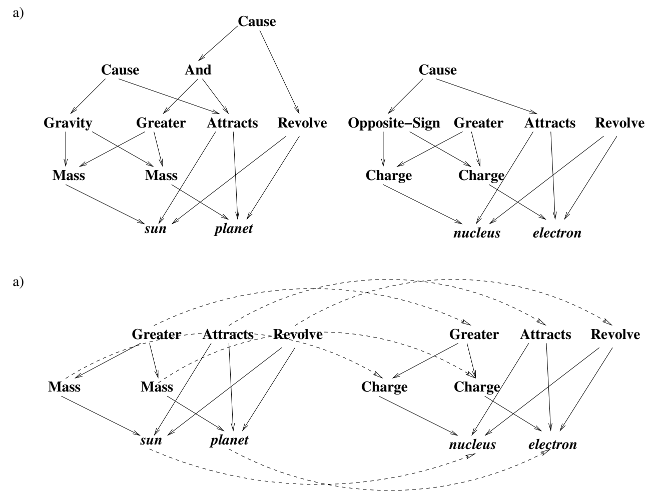

Figure 1: Analogy Derivation in Structure-Mapping Theory. (a) Two graph representations of predicate structures encoding descriptions of the solar system (left) and the Rutherford model of the atom (right). (b) An analogy between the structures in (a), where dotted arrows indicate the mappings between corresponding pairs of predicates and objects.

- Restricting the values of a set K of one or more problem-aspects (**parameters**) such that there are algorithms that are **fp-tractable** for those parameters, i.e., algorithms that run in time $f(K)n^c$ for some function f, and hence are effectively polynomial-time when those parameters are restricted (Downey & Fellows, 1999).
- Limiting the inputs to certain classes of graphs. The classes considered here are directed trees (DT), polytrees (PT) (directed acyclic graphs which remain acyclic even if the direction of their arcs is removed), polyforests (PF), and directed acyclic graphs (DAG).

Complexity Results

For reasons of space, all proofs are omitted; they can be found in Hamilton (2012). It is known that ANALOGY MAPPING is NP-hard (van Rooij et al., 2008) and hence, modulo the widely-believed conjecture that $P \neq NP$ (Fortnow, 2009), cannot be solved efficiently in general. Let $|T|$ be the size of the target graph T and d be the difference in size between T and the optimal analogy mapping.

Lemma 1 ANALOGY MAPPING *is fp-intractable for* $\{d, |T|\}$.

This shows that the properties of size and closeness that schemas possess are not sufficient on their own to guarantee efficiency. For this reason, we examine inclusion, a special case where *d* must be zero.

> (*i*, *j*)C-ANALOGY INCLUSION [(*i*, *j*)C-AI]
> *Input:* Two ordered predicate-structures *B* and *T* of class *C* with *i* and *j* roots, respectively.
> *Output:* An analogy mapping *M* between *B* and *T* such that *T* is completely mapped onto *B*, or $\perp$ if no such mapping exists.

Note that we also restrict predicate-structures to consist of predicates who arguments are ordered, *e.g.*, GREATER(X,Y). The number of roots is of special interest, as the only optimal-solution algorithm known for ANALOGY MAPPING (Falkenhainer et al., 1989) exhibits improved performance when the number of roots is restricted.

Relative to the various classes of predicate-structure graphs mentioned previously, we have the following results:

Lemma 2 *(1, 1)DT-AI can be solved in* $O(|T|)$ *time.*

Lemma 3 *(i, j)PT-AI can be solved in* $O(|T|^{1.5}|B|^2)$ *time.*

Lemma 4 *(i, j)PF-AI is NP-Hard.*

The frontier of general practicality for ANALOGY INCLUSION is thus polyforests. At this point, we must examine possible parameters to make this case solvable efficiently. Recall that *j* is the number of roots in *T* and let *f* be the maximum number of occurrence of any root predicate-type in *B* or *T*.

Lemma 5 *(i, j)PF-AI is fp-tractable for* $\{f, j\}$.

As polyforests are special cases of DAGs, this result also holds for (i, j)*DAG-AI*. Moreover, as ANALOGY INCLUSION is a special case of ANALOGY MAPPING, all fp-tractability results and most of the fp-intractability results for ANALOGY MAPPING given in van Rooij et al. (2008) and Wareham et al. (2011) hold for (i, j)*DAG-AI* as well.

Discussion

In this poster, we have shown that the frontier of polynomial-time tractability for schema-based analogy mapping is in fact lower than general DAGs and given a rough assessment of the fp-tractability options for such mapping relative to polyforests and DAGs. Much work remains to be done, both to establish the complexity of ANALOGY INCLUSION relative to all combinations of the considered parameters and to extend these results back to general schema-based analogy mapping.

There are also closely-related problems of interest. For example, it has been conjectured that ANALOGY MAPPING is easier when both *B* and *T* (rather than only *T* to *B*) are close (Gentner, 2010). The limiting case analogous to ANALOGY INCLUSION is determining the mapping between analogically identical predicate-structures. While we do not know the complexity of this problem, we do have results for a related problem, namely IDENTICAL ANALOGY, which returns "yes" if predicate-structures *B* and *T* are isomorphic, i.e. is there an analogy mapping between all of *B* and all of *T*?

Lemma 6 IDENTICAL ANALOGY *is polynomial-time equivalent to* GRAPH ISOMORPHISM.

As GRAPH ISOMORPHISM is widely believed to be polynomial-time intractable, this result provides circumstantial evidence that mutual predicate-structure closeness is not a sufficient restriction to make ANALOGY MAPPING efficiently solvable, and hence motivates the application of the methodology described here to this problem.

References

Bowdle, B., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*(1), 193–216.

Downey, R., & Fellows, M. (1999). *Parameterized Complexity*. Berlin: Springer.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The Structure-Mapping Engine: Algorithm and examples. *Artificial Intelligence*, *41*, 1–63.

Fortnow, L. (2009). The Status of the P Versus NP Problem. *Communications of the ACM*, *52*(9), 78–86.

Garey, M. R., & Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco, CA: W.H. Freeman.

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

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*(5), 752–775.

Gentner, D., Lowenstein, J., Thompson, L., & Forbus, K. D. (2009). Reviving Inert Knowledge: Analogical Abstraction Supports Relational Retrieval of Past Events. *Cognitive Science*, *33*(6), 1343–1382.

Hamilton, R. (2012). *Practical Optimal-Solution Algorithms for Schema-based Analogy Mapping*. Honours thesis, Department of Computer Science, Memorial University of Newfoundland.

van Rooij, I. (2008). The Tractable Cognition Thesis. *Cognitive Science*(32), 939–984.

van Rooij, I., Evans, P., Müller, M., Gedge, J., & Wareham, T. (2008). Identifying Sources of Intractability in Cognitive Models: An Illustration using Analogical Structure Mapping. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 915–920). Austin, TX: Cognitive Science Society.

Wareham, T., Evans, P., & van Rooij, I. (2011). What Does (and Doesn't) Make Analogical Problem Solving Easy? a Complexity-Theoretic Investigation. *Journal of Problem Solving*, *3*(2), 30–71.