Computational Complexity and KR&R: Is Polynomial Time all that Matters?[*]

Position Paper for the Workshop on Tractable Reasoning

Bernhard Nebel
German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3
W-6600 Saarbrücken, Germany
nebel@dfki.uni-sb.de

Introduction

Tractability, i.e., worst-case solvability in polynomial time, is considered as a necessary ingredient for any computational process modeling intelligent behavior from a cognitive as well as technological point of view [Levesque and Brachman, 1987; Levesque, 1988; Bylander, 1991]. Although I agree with this point of view in principle, I will argue in the following that it is often appropriate to make finer distinctions than between worst-case tractable and intractable problems:

- Classifying a problem to be *complete* for a complexity class provides us with more information about the structure of the problem and the potential sources of complexity than just classifying it as NP-hard.
- After having shown a problem to be NP-hard, it is interesting to find the precise border between tractable and intractable subproblems. Often, it turns out that "natural" restrictions lead to tractability, i.e., it is *not* necessary to develop alternative semantics or to restrict the expressiveness in a drastic way. Alternatively, such an analysis may show that the original definition of the problem is incorrect.
- Having shown polynomial time solvability does not mean that the problem can actually be solved in practice. For this purpose, efficient algorithms have to be developed. Sometimes, however, only empirical investigations can tell us which algorithm is efficient in practical cases.

In the next sections, I will justify these arguments using different examples from the literature and my own research.

Completeness Results

Often it is considered to be sufficient to prove a problem to be NP-hard and it is argued that a completeness result w.r.t. some level of the polynomial hierarchy, PSPACE, EXPTIME etc. is not relevant from a "practical" point of view. Of course, this argument is correct as far as solvability in polynomial time is concerned. However, most of the time we are interested in identifying subproblems that are tractable. For this purpose, it is helpful to find the complexity class the general problem under consideration is *complete* for.

Such results usually tell a lot about the structure of the problem and the sources of complexity. For instance, if a problem can be proven to be PSPACE-complete, such as subsumption checking in the terminological logic $\mathcal{ALC}$ [Schmidt-Schauß and Smolka, 1991], removing only one source of complexity (such as disjunction and negation or the nesting of ∀ and ∃ quantifiers) will most probably not result in tractability [Donini *et al.*, 1991]. Similar arguments hold for problems that are complete for the second level of the polynomial hierarchy such as syntax-based belief revision or default reasoning [Nebel, 1991]. Additionally, only if the potential sources of complexity of a problem have been identified, a coherent presentation of tractability results for subproblems is possible (see, e.g., [Eiter and Gottlob, 1991]), I believe.

Another interesting application of completeness results is the possibility to derive (positive or negative) results concerning the "equivalence of expressiveness." For instance, in the area of nonmonotonic logics dif-

---

[*] This work has been supported by the German Ministry for Research and Technology (BMFT) under contract ITW 8901 8.

ferent systems can be related by "translations" from one logic into another one [Konolige, 1988]. Since all propositional nonmonotonic logics (AEL, DL, circumscription) turn out to be Π_2^p-complete w.r.t. skeptical reasoning [Gottlob, 1991], it follows straightforwardly that they can be translated to each other using a polynomial transformation.

Another example concerns the relationship between abductive and consistency based diagnosis. In a recent paper, Konolige [1992, p. 257] conjectures that it is not possible to transform a general propositional causal theory in a "local way" into a form such that consistency based diagnosis gives the same results on the transformed theory as abduction on the original theory. Considering the fact that consistency based diagnosis (as described by Reiter [1987]) is NP-equivalent and abduction is Σ_2^p-equivalent, this conjecture can be confirmed (provided $\Delta_2^p \neq \Sigma_2^p$).

NP-Hardness and Tractability of Subproblems

The ultimate goal of the analysis of cognitive tasks from a computational complexity point of view is, of course, the identification of tractable methods. If a particular problem turns out to be NP-hard but the problem is apparently solved effortlessly by human beings, it is an indication that the problem formulation is too general (or simply wrong). Levesque and Brachman [1987] suggest to cope with this problem by restricting the expressiveness or by using a semantics that is weaker than the standard two-valued semantics in order to achieve tractability.

Although these are possible solutions, often "naturally occurring restrictions" are enough to make a problem tractable. Examples are the *continuous endpoint algebra* [Vilain *et al.*, 1989; Nökel, 1989], the restriction of sketch maps to maps where no river has a source on a street [Selman, 1991], and the restriction of the role-chain depth in terminologies [Nebel, 1990].

An even more drastic case is the two-level morphology. Parsing is NP-complete (in the size of the rule set) [Barton, 1986]. However, the combinatorical explosion never shows up since the number of rules responsible for the explosion is very limitied (at most 2) for each (known) natural language [Koskenniemi and Church, 1988].

Of course, it is not always possible to identify such "natural restrictions" and the only way to achieve tractability is to come up with some form of approximation or with an incomplete method. While it is not obvious how to measure the "quality" of incomplete methods, I claim that a general-purpose approximation method should at least give accurate results if it is applied to important special cases that can be solved in polynomial time. For instance, Allen's constraint propagation method to compute minimal labels for time interval networks is incomplete in general but complete for the important special case of the continuous endpoint algebra.

A negative example is Dean and Boddy's incomplete method for *temporal projection* [Dean and Boddy, 1987; Dean and Boddy, 1988]. Although the method is complete for the tractable case of totally ordered event sets, it turns out to be incomplete for another important special case that can be solved in polynomial time, namely, unconditional non-linear plans [Nebel and Bäckström, 1991]. In fact, the definition of temporal projection itself seems to be wrong since for some special cases planning appears to be easier than temporal projection. Further, the claim that temporal projection is the problem underlying *plan validation* seems not to be justified because the latter problem is tractable for unconditional, non-linear plans while the former problem is NP-hard in this case [Nebel and Bäckström, 1991; Nebel and Bäckström, 1992]. Summarizing, this seems to be an example where a complexity analysis is helpful in judging the validity of the *definition* of a problem.

Tractability versus Efficiency and the Utility of Empirical Investigations

Although proving a problem to be tractable is desirable, it does not mean that the problem can now be solved in practice. Even if a low-order polynomial algorithm has been identified, this still means that the runtime can grow faster than desired. For instance, considering terminological representation systems, the best known *classification algorithms* (for constructing the presentation of a partial order) all seem too have a worst-case and "practical case" complexity that is not better than $O(n^2)$ (n being the number of concepts) [Heinsohn *et al.*, 1992].

From a theoretical point of view, quadratic complexity is quite good. From a practical point of view, it means, however, that knowledge bases with 1000 concepts are easily dealt with while KBs that are ten times larger require a couple of hours. In order to give an impression what that means in practice, Figures 1 and 2 (taken from [Heinsohn *et al.*, 1992]) show the runtime plotted against the knowledge base size for randomly generated knowledge bases for six different terminological knowledge representation systems.

Assuming that very large knowledge bases (i.e., KBs larger than 10,000 concepts) will become relevant in the near to mid-term future, there exists clearly a need to make classification of terminological knowledge

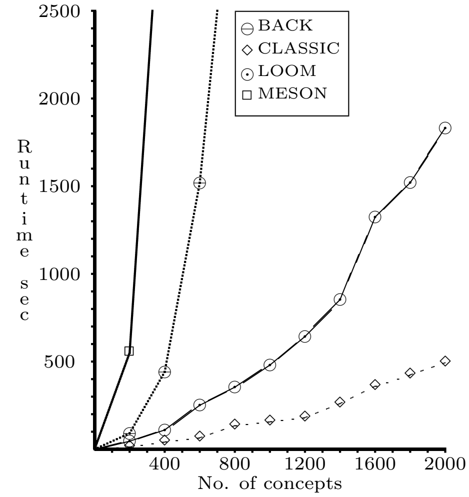

Figure 1: Runtime performance for large random KBs

bases more efficient.

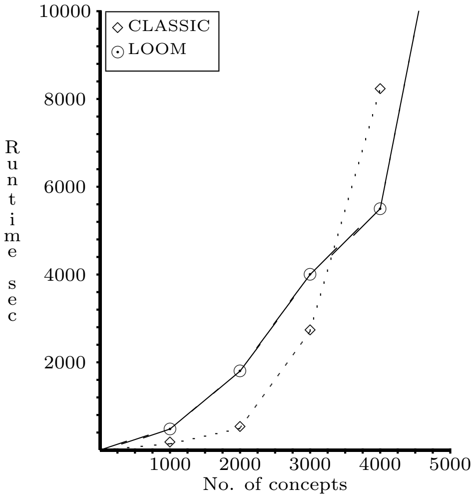

Figure 2: Runtime performance for very large random KBs

Constructing the presentation of a partial order from an underlying partial order needs $O(n^2)$ comparisons in the worst case when all elements are incomparable. So, there is no chance of finding an algorithm with a better worst case complexity. Nevertheless, knowledge bases constructed in practice seem to be better behaved. However, it is very unlikely that we will be able to find algorithms that have a *provably* better complexity in the average case since we neither know how many partial orders exist for a given cardinality [Aigner, 1988] nor do we have an idea how to describe the structure of knowledge bases occurring in practice.

The only way to proceed at this point seems to be to use empirical methods to compare different classification algorithms. In an experiment, we tested a *simple* classification method, an *enhanced* version of this method that exploits intermediate results, and a classification method using a *binary insertion strategy in a chain covering* [Baader *et al.*, 1992]. Figures 2 and 3 show the relative number of subsumption tests that were used in the simple and the enhanced method compared with the number of subsumption tests performed by the *brute force method* ($n \times (n-1)$ tests) using random KBs as test data.

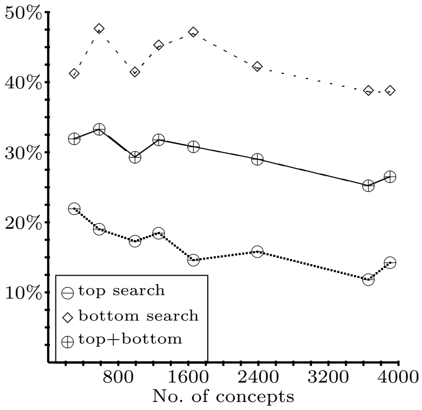

Figure 3: Number of subsumption tests for *simple method* relative to *brute force* method

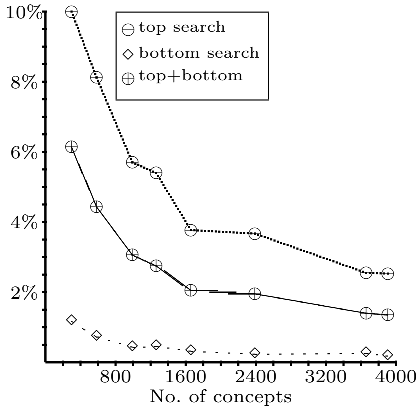

Figure 4: Number of subsumption tests for *enhanced method* relative to *brute force* method

Interestingly, it turned out that the algorithm we conjectured to be the fastest one—the algorithm using binary insertion—is indeed very efficient on randomly generated partial orders but slower than the enhanced method on knowledge bases occurring in practice. This result shows that empirical investigations are indeed necessary. Another interesting result of our investigation was that the optimizations in the second method are effective only for some knowledge bases, namely, those that where designed for linguistic purposes. We do not have an explanation for this somehow funny behavior but conjecture that the reason for this behavior has to do with the fact that the linguistic knowledge bases contain more *implicit* relations between concepts than the other KBs.

References

AAAI-92. *Proceedings of the 10th National Conference of the American Association for Artificial Intelligence*, San Jose, CA, July 1992. MIT Press.

Martin Aigner. *Combinatorical Search.* Teubner, Stuttgart, Germany, 1988.

James A. Allen, Richard Fikes, and Erik Sandewall, editors. *Principles of Knowledge Representation and Reasoning: Proceedings of the 2nd International Conference*, Cambridge, MA, April 1991. Morgan Kaufmann.

Franz Baader, Bernhard Nebel, Hans-Jürgen Profitlich, and Enrico Franconi. An empirical analysis of optimization techniques for terminological representation systems or “making KRIS get a move on”. DFKI Research Report, German Research Center for Artificial Intelligence (DFKI), Saarbrücken, 1992. In preparation.

G. E. Barton. Computational complexity in two-level morphology. In *Proceedings of the 24th Annual Meeting of the ACL*, pages 53–59, New York, NY, 1986.

Tom Bylander. Tractability and artificial intelligence. *Journal of Experimental and Theoretical Artificial Intelligence*, 3:171–178, 1991.

Thomas L. Dean and Mark Boddy. Incremental causal reasoning. In *Proceedings of the 6th National Conference of the American Association for Artificial Intelligence*, pages 196–201, Seattle, WA, July 1987.

Thomas L. Dean and Mark Boddy. Reasoning about partially ordered events. *Artificial Intelligence*, 36(3):375–400, October 1988.

Francesco M. Donini, Maurizio Lenzerini, Daniele Nardi, and Werner Nutt. The complexity of concept languages. In Allen et al. [1991], pages 151–162.

Thomas Eiter and Georg Gottlob. On the complexity of propositional knowledge base revision, updates, and counterfactuals. Technical Report CD-TR-91/23, Christian Doppler Labor, Technische Universität Wien, Wien, Austria, 1991.

Georg Gottlob. Complexity results for nonmonotonic logics. Technical Report CD-TR 91/24, Christian Doppler Labor, Technische Universität Wien, Wien, Austria, 1991.

Jochen Heinsohn, Daniel Kudenko, Bernhard Nebel, and Hans-Jürgen Profitlich. An empirical analysis of terminological representation systems. In AAAI-92 [1992].

Kurt Konolige. On the relation between default and autoepistemic logic. *Artificial Intelligence*, 35(3):343–382, July 1988.

Kurt Konolige. Abduction versus closure in causal theories. *Artificial Intelligence*, 53(2-3):255–272, February 1992.

Kimmo Koskenniemi and Kenneth Ward Church. Complexity, two-level morphology and finish. In *Proceedings of the 12th International Conference on Computational Linguistics*, pages 335–340, Budapest, Hungary, 1988.

Hector J. Levesque and Ronald J. Brachman. Expressiveness and tractability in knowledge representation and reasoning. *Computational Intelligence*, 3:78–93, 1987.

Hector J. Levesque. Logic and the complexity of reasoning. *Journal of Philosophical Logic*, 17:355–389, 1988.

Bernhard Nebel and Christer Bäckström. On the computational complexity of temporal projection and some related problems. DFKI Research Report RR-91-34, German Research Center for Artificial Intelligence (DFKI), Saarbrücken, 1991. Also published as Research Report LiTH-IDA-R-91-34, Department of Computer and Information Science, Linköping University, Linköping, Sweden.

Bernhard Nebel and Christer Bäckström. On the computational complexity of temporal projection and plan validation. In AAAI-92 [1992]. To appear.

Bernhard Nebel. Terminological reasoning is inherently intractable. *Artificial Intelligence*, 43:235–249, 1990.

Bernhard Nebel. Belief revision and default reasoning: Syntax-based approaches. In Allen et al. [1991], pages 417–428.

Klaus Nökel. Convex relations between time intervals. In J. Rettie and K. Leidlmair, editors, *Proceedings der 5. Österreichischen Artificial Intelligence-Tagung,* pages 298–302. Springer-Verlag, Berlin, Heidelberg, New York, 1989.

Raymond Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32(1):57–95, April 1987.

Manfred Schmidt-Schauß and Gert Smolka. Attributive concept descriptions with complements. *Artificial Intelligence*, 48:1–26, 1991.

Bart Selman. Sketch map interpretation: a complexity analysis. AT&T Bell Labs., Murray Hill, NJ, 1991.

Marc B. Vilain, Henry A. Kautz, and Peter G. van Beek. Constraint propagation algorithms for temporal reasoning: A revised report. In D. S. Weld and J. de Kleer, editors, *Readings in Qualitative Reasoning about Physical Systems*, pages 373–381. Morgan Kaufmann, San Mateo, CA, 1989.