Parameterized Complexity Analysis in Cognitive Science

Introduction

Many computational problems arising within Cognitive Science are known to be intractable, *i.e.*, there are no algorithms that can compute exact / optimal solutions for all instances of those problems within time polynomially bounded in the instance size. Two sources of such problems are:

1. **Computational-level theories of cognitive abilities**: A cognitive ability such as belief maintenance, visual search, analogy derivation, or goal inference from observed actions typically has one or more associated computational-level theories which describe in general terms how that ability is implemented in the brain. Each such theory in turn has one or more associated computational problems which encode the mental representations and mechanisms proposed by that theory. These theories guide both pure research into cognition and implementations of cognitive abilities in practical Artificial Intelligence (AI) software systems.

2. **Computational analyses of large cognitive datasets**: The last 25 years has seen a dramatic increase in the number and size of multi-media databases of observations of various cognitive abilities, *e.g.*, the Phon child language acquisition database [19]. In addition to regular database operations such as lookup of known items, analyses of these datasets should also include derivation of patterns in the data such as longest common patterns (if the data is stored as temporal sequences of behavior or performance) or clusterings of observations or observation-sequences by similarity. Such analyses can be used both to evaluate existing cognitive theories or hypotheses and to develop new ones.

For problems that occur in software systems, polynomial-time solvability is necessary to ensure reasonable system runtimes relative to large datasets. For problems associated with cognitive theories, given that computational-level cognitive theories describe in a broad sense how cognitive abilities may be implemented in the brain and human cognitive abilities run quickly in practice, polynomial-time solvability seems a natural necessary (though not sufficient) constraint for assessing the plausibility of these theories [1, 21, 23].

Following common practice in Computer Science, the typical approach to dealing with such intractable problems in Cognitive Science is to invoke polynomial-time approximation or heuristic algorithms, either as implementations in software systems or as proposed mechanisms in the underlying cognitive theories. This approach runs the risks of impairing the solution-quality performance of implemented systems and, perhaps more insidiously, introducing logical inconsistencies into or even misleading the process of reasoning about cognition [21, 24]. Moreover, this approach may be unnecessary, given the many natural restrictions on cognitive inputs and outputs and the possibility that some existing exponential-time exact / optimal algorithms may run in effectively polynomial-time under these restrictions.

The research in this proposal will use techniques from parameterized complexity theory [3] to explore the exponential-time algorithmic possibilities for the exact / optimal solution of various problems arising in Cognitive Science. This research will focus on the following problem-areas:

- **Analogy derivation**: Given two conceptual structures, analogy derivation looks for the best possible analogy between these structures relative to some criterion, where an analogy is essentially a common subgraph of the given structures. Analogy derivation (especially as formulated within the structure-mapping theory (SMT) proposed by Gentner [10]) plays a central role in Cognitive Science, and underlies a number of important processes including conceptual structure re-representation [28], exemplar retrieval [7], and exemplar generalization [14], many of which have applications in AI systems.

- **Common subsequence derivation and sequence alignment**: Given a set of two or more sequences, one can look for the largest common subsequence shared by these sequences or the best possible alignment of corresponding portions of these sequences. As common subsequences can be used to construct alignments and alignments can be used to construct common subsequences, these problems are very closely related. If temporal cognitive observation data is interpreted as sequences, common subsequences and alignments can be used as summaries of common temporal development patterns and variations from these patterns, respectively. Though such common subsequences and alignments have a long-standing history of use within molecular biology and speech processing [20], they have been little-used in Cognitive Science (particularly in analyses of > 2 given sequences).

Both analogy derivation and many formulations of common subsequences and alignment over multiple sequences are known to be NP-hard [16, 22, 25, 26] and hence are considered intractable. There are many techniques [3, 18] for establishing which subsets of aspects of a problem have fixed-parameter tractable (FPT) algorithms, *i.e.*, algorithms whose exponential behavior is phrased purely in terms of those aspects and hence run in polynomial-time if these aspects are of fixed or constant value. The results derived by a systematic application of these techniques to all subsets of a selected set of limited-value aspects of a problem can be organized into an "intractability map" [27] that summarizes which subsets have associated FPT algorithms and, equally importantly, which subsets do not. Those subsets yielding FPT algorithms may then be considered either worth investigating for better FPT algorithms or (in the case of problems associated with cognitive theories) sources of intractability in those theories and hence associated with mechanisms whose revision may yield theory tractability [22, 23]

Recent Progress

The methodological foundations for using parameterized complexity to analyze cognitive theories were laid out independently by Dr. Iris van Rooij and myself in our doctoral work on belief revision and phonological processing problems, respectively (see summaries in [21, 23]). Since we started collaborating in 2006, we have worked on various problems of interest in Cognitive Science [13, 17, 24], perhaps most notably giving the first formal analysis of the sources of polynomial-time intractability in analogy derivation under SMT [22].

I have done little work on common subsequence or alignment problems recently, outside of a summary of known FPT algorithms and parameterized intractability results for the LONGEST COMMON SUBSEQUENCE and COMMON APPROXIMATE SUBSTRING problems in [5]. However, my previous experience doing complexity analyses of common substring and subsequence [2, 4] and common dataset pattern [12] problems should be of use here, in conjunction with my ongoing algorithm design work on Phon [9, 15, 19].

Objectives

My short-term, *i.e.*, < 1.5 years, objectives are:

1. Finish the intractability map for analogy derivation under SMT started in [22]; formulate and perform complexity analyses of analogy-related problems, *e.g.*, re-representation, exemplar generalization, exemplar retrieval.

2. Formulate and perform initial complexity analysis of Phon-useful variants of common subsequence and sequence alignment problems.

My medium-term, *i.e.*, 1.5 to 3 year, objectives are:

1. Explore possible revisions to analogy derivation under SMT and related problems relative to the derived intractability maps; develop, implement, and test the best possible FPT algorithms for these problems.

2. Finish intractability map for Phon-useful variants of common subsequence and sequence alignment problems; develop best possible FPT algorithms for these problems.

My long-term, *i.e.*, 3 to 5+ year, objectives are:

1. Perform parameterized analyses of other analogy-related problems, *e.g.*, metaphor processing, as well as problems associated with other cognitive abilities, *e.g.*, skill acquisition, goal inference from observed actions.

2. Develop, implement, and test the best possible common subsequence and sequence alignment algorithms (FPT, polynomial-time approximation, or heuristic) for implementation in Phon.

Methods and Proposed Approach

The process for creating an intractability map for a computational problem is as follow:

1. Examine the literature to find known algorithms and intractability results for the problem.

2. Examine both the literature and applications of the problem for useful aspects, *e.g.*, aspects that are of bounded size or value in practice.

3. Perform a systematic parameterized analysis relative to the aspects selected in (2).

With respect to complexity analysis of cognitive abilities, where the primary intent is to revise cognitive theories to make them tractable in the interests of cognitive plausibility [21, 22, 23], the core process above may be iteratively repeated relative to problems associated with revised cognitive theories and (possibly augmented) sets of problem aspects. If the intent is to derive practical algorithms (as is the case for application of cognitive-ability algorithms to AI systems and the analysis of large cognitive observation datasets), two more steps follow the process above:

4. Derive the best possible FPT algorithms for the problem relative to the fixed-parameter tractable aspect-subsets discovered in (3).

5. Implement the algorithms derived in (4) and test them on simulated and real datasets to establish practical vs. theoretical performance.

If *any* practical (instead of just practical FPT) algorithms are acceptable, as may be the case in analyzing large cognitive observation datasets, steps (4) and (5) above may also include derivation, implementation, and testing of polynomial-time approximation and heuristic algorithms.

Given the breadth of expertise required to carry out research in the manner described above relative to the problems considered in this proposal, appropriate research collaborators are essential. Thus, a key part of this proposed research will be my ongoing collaborations with Drs. Iris van Rooij (given her expertise in psychology, cognitive science, and parameterized complexity analysis) and Patricia Evans (given her expertise in parameterized complexity analysis and algorithm development with respect to common substring, subsequence, and subgraph problems).

Pertinent Literature

All parameterized complexity analyses of cognitive theories to date have been done by Iris van Rooij, myself, and our co-authors. Though there is an extensive psychology literature on analogy and analogy-related processes as well as these processes under SMT, there are very few formal statements of the computational-level problems associated with these processes (the most notable exceptions perhaps being [6, 10, 28]). The only published optimal-solution algorithm for an analogy-based problem under SMT is that for analogy derivation as implemented in the original version of the Structure Mapping Engine (SME) [6], and the only complexity analyses of the intractability of such a problem (again, analogy derivation) are those in [25, 22]. As analogy derivation under SMT is essentially a very restricted version of the common subgraph problem on directed acyclic graphs (DAGs), it is conceivable that potentially applicable exponential-time algorithms have already been derived; however, as a preliminary literature search suggests that research on common subgraph problems is scattered over a wide range of application areas and subdisciplines within Computer Science, it is not yet obvious where such algorithms may be. The Bioinformatics literature might be particularly promising, as biological structures and common subgraph problems on these structures are frequently phrased in terms of restricted versions of DAGs such as directed trees and arc-annotated sequences.

There is an extensive literature on both common subsequence and sequence alignment problems in Computer Science and in particular in Bioinformatics (see [5, 11] and references). Though a number of optimal-solution algorithms have been described for these problems, it is not yet obvious which of these may be applicable or relevant to cognitive observation databases such as Phon.

Anticipated Significance of Work

The results of the proposed complexity analyses of cognitive theories will be used in evaluating existing and deriving new computational- level cognitive theories of the abilities examined (along the lines sketched in [21, 22, 23]), which in turn should lead to AI implementations of these abilities with better solution-quality (and possibly run-time) performance. The results of the proposed complexity analyses of methods for analyzing large cognitive observation datasets will be of immediate interest to both the developers and users of Phon, and should provide a general framework for the analysis of large cognitive observation datasets.

Contribution to the Training of Personnel

Over the course of the next five years, I expect to supervise four to six graduate students, three to five undergraduate project (B.Sc. Honours) students, and four to eight undergraduate student research assistant. I hope to establish a “personnel pipeline” leading from undergraduate student to graduate student, so that experience in the various projects described above can be acquired gradually (and thus more thoroughly) over several years instead of the four months and two years typically allotted to undergraduate projects and M.Sc. graduate degrees, respectively. Given my experience with interdisciplinary research between computer science, biology and linguistics (and most lately, cognitive science), I can also give these students a good grounding in and an appreciation of the joys of interdisciplinary research.

My focus on Masters-level and undergraduate students may seem unusual in an interdisciplinary project such as that proposed here, which seems to require doctoral or post-doctoral qualifications. My own experience training students over the last 6 years (in particular with Chaytor, Gedge, Hamilton, and Uddin) has demonstrated to me that any motivated and bright student (regardless of degree-level) can make valuable and publishable contributions relatively quickly in an appropriately-chosen research area, especially if they receive frequent mentoring at the beginning of their research. I enjoy working with students and feel that I can give them the required mentoring; moreover, given the novelty of (and hence potential for) parameterized applications in Cognitive Science and AI as well as the scarcity of even basic formal complexity analyses in Cognitive Science and AI, I believe that the research described in this proposal is a most appropriate research area at this time.

References

[1] Barton, G.E., Berwick, R.C., and Ristad, E.S. (1987) *Computational Complexity and Natural Language*. MIT Press.

[2] Bodlaender, H., Downey, R.G., Fellows, M.R., and Wareham, H.T. (1995) The Parameterized Complexity of Sequence Alignment and Consensus. *Theor. Comp. Sci.*, 147(1–2), 31–54.

[3] Downey, R.G. and Fellows, M.R. (1999) *Parameterized Complexity*. Springer.

[4] Evans, P.A., Smith, A.D., and Wareham, H.T. (2003) On the Complexity of Finding Common Approximate Substrings. *Theor. Comp. Sci.*, 306(1–3), 407–430.

[5] Evans, P.A. and Wareham, H.T. (2009) Efficient Restricted-Case Algorithms for Problems in Computational Biology. To appear in *Algorithms in Computational Molecular Biology*. Wiley.

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

[7] Forbus, K.D., Gentner, D., and Law, K. (1994) MAC / FAC: A Model of Similarity-based Retrieval. *Cognitive Science*, 19, 141–205.

[8] Garey, M.R. and Johnson, D.S. (1979) *Computers and Intractability* W.H. Freeman.

[9] Gedge, J., Hedlund, G.J., Rose, Y., and Wareham, T. (2007) Natural Language Process Detection: From Conception to Implementation. In *NECEC 2007*.

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

[11] Gusfield, D. (1997) *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press.

[12] Hamilton, M., Chaytor, R., and Wareham, T. (2006) The Parameterized Complexity of Enumerating Frequent Itemsets. In *IWPEC 2006*. LNCS no. 4196. Springer. 227–238.

[13] Hamilton, M., Müller, M., van Rooij, I., and Wareham, T. (2007) Approximating Solution Structure. In *Structure Theory and FPT Algorithmics for Graphs, Digraphs, and Hypergraphs.* Dagstuhl Seminar Proceedings no. 07281. IBFI.

[14] Kuehne, S.E., Forbus, K.D., Gentner, D., and Quinn, B. (2000) SEQL: Category learning as progressive abstraction using structure mapping. In *CogSci 2000*. 770–775.

[15] Maddocks, K., Hedlund, G., Rose, Y., and Wareham, T. (2005) Pairwise Natural Language Syllabic Alignment: From Conception to Implementation. In *NECEC 2005*.

[16] Maier, D. (1978) The complexity of some problems on subsequences and supersequences. *Journal of the ACM*, 25, 322–336.

[17] Müller, M., van Rooij, I., and Wareham, T. (2009) Similarity as Tractable Transformation. In *CogSci 2009*. 49–55.

[18] Niedermeier, R. (2006) *Invitation to Fixed-Parameter Algorithms*. Oxford University Press.

[19] Rose, Y., Hedlund, G.J., Byrne, R., Wareham, T. and MacWhinney, B. (2007) Phon 1.2: A Computational Basis for Phonological Database Elaboration and Model Testing. In *Proc. Workshop on Cognitive Aspects of Computational Language Acquisition (ACL 2007).* 17–24.

[20] Sankoff, D. and Kruskal, J.B. (eds.) (1983) *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. Addison-Wesley.

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

[22] van Rooij, I., Evans, P., Müller, M., Gedge, J., and Wareham, T. (2008) Identifying Sources of Intractability in Cognitive Models: An Illustration using Analogical Structure Mapping. In *CogSci 2008*. 915–920.

[23] van Rooij, I. and Wareham, T. (2008) Parameterized Complexity in Cognitive Modeling: Foundations, Applications, and Opportunities. *Computer Journal*, **51(3)**, 385–404.

[24] van Rooij, I., Wright, C., and Wareham, T. (2009) Intractability and the Use of Heuristics in Psychological Explanations. Submitted to *Philosophy of Science*.

[25] Veale, T. and Keane, M.T. (1997) The Competence of Sub-Optimal Theories of Structure Mapping on Hard Analogies. In *IJCAI 1997*. Morgan Kaufmann. 232–237.

[26] Wang, L. and Jiang, T. (1994) On the complexity of multiple sequence alignment. *J. Comput. Biol.*, 1, 337–348.

[27] Wareham, H.T. (1999) *Systematic Parameterized Complexity Analysis in Computational Phonology.* Ph.D. thesis, Department of Computer Science, University of Victoria.

[28] Yan, J., Forbus, K.D., and Gentner, D. (2003) A Theory of rerepresentation in analogical matching. In *CogSci 2003*. 1265–1270.