How Many Dimensions in Co-Optimization?

Wojciech Jaśkowski
Institute of Computing Science
Poznan University of Technology
Piotrowo 2, 60965 Poznań, Poland
wjaskowski@cs.put.poznan.pl

Krzysztof Krawiec
Institute of Computing Science
Poznan University of Technology
Piotrowo 2, 60965 Poznań, Poland
kkrawiec@cs.put.poznan.pl

ABSTRACT

Co-optimization test-based problems is a class of tasks approached typically with coevolutionary algorithms. It was recently shown that such problems exhibit underlying objectives that form internal problem structure, which can be extracted and analyzed in order to drive the search or design better algorithms. The number of underlying objectives is the dimension of the problem, which is of great importance, since it may be a predictor of problem's difficulty. In this paper, we estimate the number of dimensions for Tic Tac Toe and the Density Classification Task.

Categories and Subject Descriptors

I.2.8 [**Problem Solving, Search**]: Heuristic methods

General Terms

Algorithms

Keywords

Test-Based Problem, Coevolution, Co-optimization, Games, Dimensionality, Density Classification Task, Tic Tac Toe

1. INTRODUCTION

A co-optimization test-based problem consists of a set of candidate solutions (candidates) S, a set of tests T and an interaction function $G : S \times T \rightarrow V$. The goal of the problem is to find the best candidate in S with respect to some preference relation. T is usually large, thus computing the vector of all interaction outcomes for a given candidate is infeasible, making the problem hard to tackle.

A promising direction of research in this field involves extracting internal multi-objective structure of a test-based problem [5], which could help designing better coevolutionary algorithms [2] and examining problem properties [3]. To the best of our knowledge, all of previous research in extracting internal problem structure was done either on artificial number games [1] or a very small version of Nim [3]. Here we examine more complex test-based problems and try to estimate how many dimensions such problems have.

To this aim we consider Bucci's definition of coordinate system [1], which models the internal structure of a problem. Using GreedyCover heuristic [4], we estimate the number of dimensions of two problems: Tic Tac Toe and the Density Classification Task.

GECCO'11, July 12–16, 2011, Dublin, Ireland.
ACM 978-1-4503-0690-4/11/07.

2. METHODS

In the context of test-based problems, a coordinate system [1] is a formal concept revealing the internal problem structure by enabling candidates and tests to be embedded in a multidimensional space. It is defined for problems with binary outcome $V = \{0, 1\}$, thus a candidate can only fail or solve a test. An axis of a coordinate system (interpreted as an underlying objective) consists of tests ordered by the dominance relation. A candidate is positioned in the space spanned by such axes, so that its dominance cone contains all tests it solves. A coordinate system is correct if it preserves all dominance relations between candidates defined by the interaction function G [4]. Of particular interest are correct coordinate systems of minimal number of axes, which give rise to the notion of dimension of test-based problem.

Given S, T and G, our GreedyCover heuristic [4] finds a correct coordinate system.

3. EXPERIMENTS

It was shown [4] that the dimension of a random test-based problem (a random $n \times n$ payoff matrix) grows with a logarithm of the problem size n (see Fig. 1), which indicates that the compression provided by underlying objectives may be exponential with respect to $|T|$. In order to we verify how this result generalizes to more complex test-based problems, we performed experiments on two other problems: Tic Tac Toe and Density Classification Task.

Tic Tac Toe (TTT).

Player's X strategy is identified with a candidate while player's O with a test. The interaction function returns 0 for the candidate passing the test (win, draw) and 1 for failing it (defeat). Strategies are encoded directly. TTT has 765 different board states (excluding rotated and reflected ones), of which 627 are not final. In 338 states it is X's turn and in 289 O is to move. Thus, 338 and 289 genes define X and O strategies, respectively. A gene encodes the move to play in one state. The total number of strategies is roughly 3.47×10^{162} for player X and 2.82×10^{142} for player O.

Density Classification Task (DCT).

The goal in this problem is to find a one-dimensional cellular automaton rules that perform majority voting. We

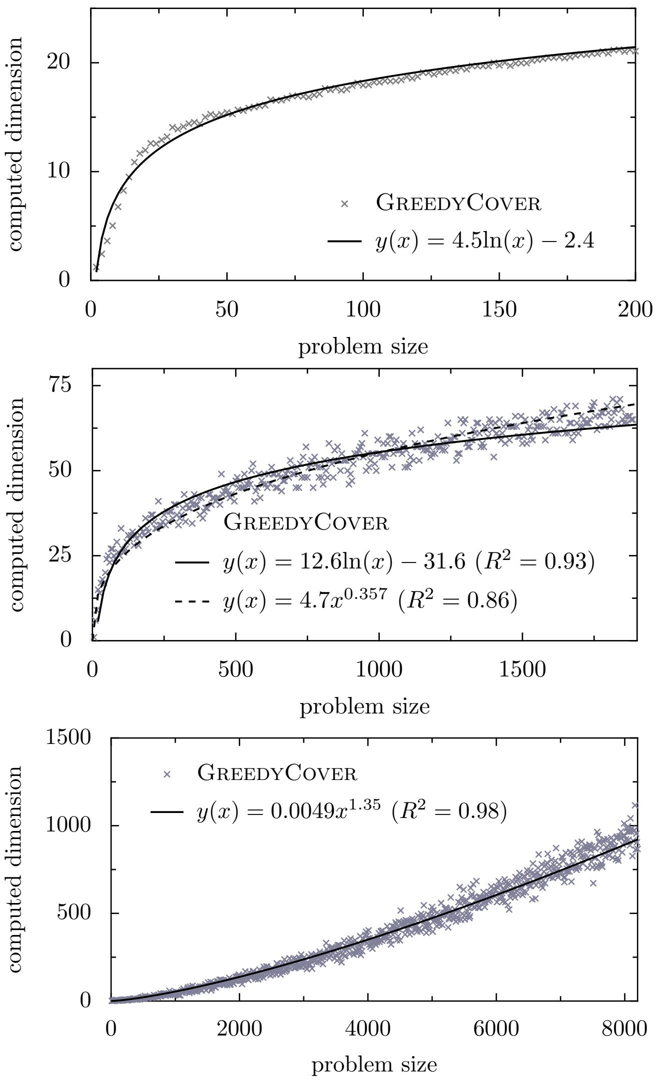

Figure 1: Dimension trends for random test-based problem (upper), TTT (middle), DCT (bottom)

consider an instance of DCT with an initial configuration size $n = 31$ and rule radius $r = 2$.

4. RESULTS

The number of strategies in both TTT and DCT is too high to analyze the whole payoff matrix, thus we randomly sampled strategies from the entire strategy space. We were able to consider payoff matrices from 5×5 to 1885×1885 for TTT, and 10×10 to 8200×8200 for DCT. For each of them, we used GreedyCover to compute the number of dimensions.

Figure 1 shows the results of this procedure for TTT and DCT. For TTT, two standard trend curves fit our data: a power function $y = 4.7x^{0.357}$ with $\mathrm{R}^2 = 0.86$; and a logarithmic one $y = 12.62\ln(x) - 31.6$ with $\mathrm{R}^2 = 0.93$. The difference between R^2 coefficients is small, so there is not enough evidence to confidently claim that the logarithmic curve describes the dimension trend better. However, as GreedyCover is a heuristic, the bigger the problem, the more it overestimates the dimension, which may be an argument in favor of the logarithmic curve.

We use the trend to estimate *the upper bound of the number of dimensions* of TTT. Since we are unable to unambiguously determine the trend function, we may make claims only conditionally. If we assume that the power function correctly determines the dimension trend, then extrapolating it to 3.47×10^{162} strategies yields roughly 5.0×10^{58} dimensions, however assuming that the logarithm function provides better fit, we get approx. only 2038 dimensions.

For DCT, the data are best modeled by a power function $y(x) = 0.0047x^{1.35}$ with $R^2 = 0.98$.

5. DISCUSSION AND CONCLUSIONS

Similarly to the random problem, the dimension trend for TTT may be a logarithmic function, but it is clearly a power function for DCT. We conclude that the relation between the dimension of a problem sample and the size of the sample is highly problem-dependent.

In multi-objective optimization, generally, the more objectives, the harder the problem. The same may hold for test-based problems, as they can be interpreted as special cases of multi-objective optimization problem. Hence, random sampling in order to estimate the dimension trend may be a tool of practical interest. In this light, TTT turns out to be much easier than DCT.

Apart from that, we would like to point out that the popularly used 2 or 3-dimensional number games (compare-on-all and compare-on-one) are not representative as test-beds for co-optimization problems, since, as we have seen for TTT or DCT, the number of dimensions of real problems are much higher. Thus, results on such simple number games could not be generalized and can be misleading.

6. ACKNOWLEDGMENT

This research has been supported by the Ministry of Science and Higher Education grant # N N519 441939.

7. REFERENCES

[1] A. Bucci, J. B. Pollack, and E. de Jong. Automated extraction of problem structure. In K. Deb et al., editor, *GECCO'04*, volume 3102 of *LNCS*, pages 501–512, Seattle, WA, USA, 26-30 June 2004. Springer-Verlag.

[2] E. de Jong and A. Bucci. DECA: dimension extracting coevolutionary algorithm. In M. Cattolico et al., editor, *GECCO'06*, pages 313–320, Seattle, Washington, USA, 2006. ACM Press.

[3] E. de Jong and A. Bucci. Objective Set Compression. Test-Based Problems and Multiobjective Optimization. In J. Knowles et al., editor, *Multiobjective Problem Solving from Nature: From Concepts to Applications*, pages 357–376. Springer, Berlin, 2008.

[4] W. Jaśkowski and K. Krawiec. Formal analysis and algorithms for extracting coordinate systems of games. In *IEEE Symposium on Computational Intelligence and Games*, pages 201–208, Milano, Italy, 2009.

[5] E. Popovici, A. Bucci, P. Wiegand, and E. de Jong. *Handbook of Natural Computing*, chapter Coevolutionary Principles. Springer-Verlag, 2011.