Artificial Intelligence 228 (2015) 129–152

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Confidence-based reasoning in stochastic constraint programming ☆

Roberto Rossi [a,*], Brahim Hnich [b], S. Armagan Tarim [c,d,1], Steven Prestwich [d,1]

[a] Business School, University of Edinburgh, United Kingdom
[b] Department of Computer Science, Taif University, Taif, Saudi Arabia
[c] Department of Management, Cankaya University, Ankara, Turkey
[d] Insight Centre for Data Analytics, University College Cork, Ireland

ARTICLE INFO

Article history:
Received 7 November 2014
Received in revised form 5 July 2015
Accepted 8 July 2015
Available online 15 July 2015

Keywords:
Confidence-based reasoning
Stochastic constraint programming
Sampled SCSP
(α, ϑ)-solution
(α, ϑ)-solution set
Confidence interval analysis
Global chance constraint

ABSTRACT

In this work we introduce a novel approach, based on sampling, for finding assignments that are likely to be solutions to stochastic constraint satisfaction problems and constraint optimisation problems. Our approach reduces the size of the original problem being analysed; by solving this reduced problem, with a given confidence probability, we obtain assignments that satisfy the chance constraints in the original model within prescribed error tolerance thresholds. To achieve this, we blend concepts from stochastic constraint programming and statistics. We discuss both exact and approximate variants of our method. The framework we introduce can be immediately employed in concert with existing approaches for solving stochastic constraint programs. A thorough computational study on a number of stochastic combinatorial optimisation problems demonstrates the effectiveness of our approach.

1. Introduction

The stochastic constraint satisfaction/optimisation framework introduced in [2,3] constitutes an expressive declarative formalism for modelling problems of decision making under uncertainty. A stochastic constraint satisfaction problem (SCSP), alongside decision variables, features *random variables*, which follow some probability distribution and can be used to model uncertainty. Relationships over subsets of random and decision variables can be expressed in a declarative manner via *stochastic constraints*. The fact that a given relationship over subsets of random and decision variables should be satisfied according to a prescribed probability can be expressed by means of *chance constraints*. Finally, since problems of decision making under uncertainly are sequential in nature, the modeller can define a *stage structure*, that is a sequence of *decision stages*, in each of which a subset of all possible decisions are taken and a subset of all possible random variables are observed. A solution to an SCSP can be represented in general by means of a *policy tree*, which records feasible or optimal decisions associated with each possible set of random variable realisations.

☆ This work is an extended version of [1].

* Corresponding author at: Business School, University of Edinburgh, 29 Buccleuch place, EH8 9JS, Edinburgh, UK. Tel.: +44 (0)131 6515239; fax: +44 (0)131 650 8077.

E-mail addresses: roberto.rossi@ed.ac.uk (R. Rossi), hnich.brahim@gmail.com (B. Hnich), armtar@yahoo.com (S.A. Tarim), s.prestwich@cs.ucc.ie (S. Prestwich).

1 This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289.

http://dx.doi.org/10.1016/j.artint.2015.07.004

As shown in [3, Theorem 1], solving SCSPs is a computationally hard task. Even trivial instances with a dozen of decision and random variables require a computational effort out of reach even for the most advanced hardware/software combination. This is due to the fact that the size of the policy tree grows exponentially in the number of random variables in the model and in the size of their support. Furthermore, a major limitation of all existing complete SCSPs solution methods, such as [3] and [4], is the fact that they assume the support of random variables is *finite*, otherwise a solution cannot be expressed as a finite policy tree. In practice, however, it is often the case that random variables either range over continuous supports or have a very large number (possibly infinite) of values in their domain. To date, no general purpose method exists for solving large-scale SCSPs, or SCSPs featuring random variables with continuous or discrete infinite support; for the sake of brevity we shall name this latter class of SCSPs "infinite SCSPs."

The main contribution of this paper is to propose a framework for solving large-scale or infinite SCSPs. More specifically, we argue that in solving large-scale or infinite SCSPs, one should not consider the ultimate feasible/optimal solution, which in some cases may even be impossible to represent; rather, the decision maker should aim for a solution that she "sufficiently trusts," which she may claim to be optimal or feasible with a given confidence level, and for which a certain degree of error may be tolerated. In order to obtain such a solution, the decision maker should only look at a possibly limited number of samples drawn from the random variables in the model. In other words, she should try to "estimate" the quality of this solution.

Our approach has several analogies with established techniques in statistics. When a survey is conducted on a sample population — e.g. an electoral poll — a statistician typically associates a certain confidence level with the results obtained from the chosen sample population. For instance, one may claim that there is a 90% chance that the actual mean being estimated is within a given interval. We argue that the very same approach may be adopted in stochastic decision making. If the infinite or large-scale m-stage SCSP does not admit any closed form solution and is complex enough to rule out any chance of obtaining an exact solution, we suggest that — as is done in statistics — one may introduce a confidence level α and a tolerated estimation error $\pm\vartheta$. The decision maker, instead of looking for an exact solution, may then aim to "estimate" — according to the chosen α and ϑ — whether the actual satisfaction probability guaranteed by an assignment is greater than or equal to the given target value for each of the chance constraints in the model. By choosing given values for α and ϑ the set of solutions may vary. For this reason we will introduce a new notion of solution that is parameterised by these two parameters and that we call an (α, ϑ)-solution. Intuitively, as α tends to 1 and ϑ tends to 0 the set of (α, ϑ)-solutions will converge to the set of actual solutions to the original stochastic constraint satisfaction problem, which we therefore rename $(1, 0)$-solutions. One should note that an approach of this kind has been recently advocated in [5, Chap. 4].

In this work, we make the following contributions to the stochastic constraint programming literature:

- we discuss how to obtain compact instances of infinite or large-scale stochastic constraint programs via sampling: we call these instances "sampled SCSPs;"
- we introduce the concepts of (α, ϑ)-solution and of (α, ϑ)-solution set; and show how to compute a priori the minimum sample size that guarantees the attainment of such classes of solutions;
- we show how the above tools can be employed in order to find approximate solutions to infinite or large-scale stochastic constraint satisfaction/optimisation problems that cannot be solved by existing exact approaches in the stochastic constraint programming literature;
- we conduct a thorough computational study on three well-known stochastic combinatorial problems to validate our theoretical framework and assess its effectiveness, efficiency, and scalability.

This work is structured as follows: in Section 2 we introduce the relevant formal background in constraint programming, stochastic constraint programming, and confidence interval analysis; in Section 3 we introduce sampled SCSPs; in Section 4 we discuss properties of the solutions of sampled SCSPs and formally introduce (α, ϑ)-solutions; in Section 5 we introduce (α, ϑ)-solution sets; in Section 6 we extend our discussion to stochastic constraint optimisation problems; in Section 7 we discuss connections with established techniques in statistics; in Section 8 we present our computational study; in Section 9 we discuss related works; finally, in Section 10 we draw conclusions and discuss future research directions.

2. Formal background

We now introduce the relevant background in constraint programming, stochastic constraint programming, and confidence interval analysis.

2.1. Constraint programming

A Constraint Satisfaction Problem (CSP) [6] consists of a set of decision variables, each with a finite domain of values, and a set of constraints specifying allowed combinations of values for some variables. A *solution* to a CSP is an assignment of variables to values in their respective domains such that all of the constraints are satisfied. Constraint solvers typically explore partial assignments enforcing a local consistency property. A constraint c is *generalised arc consistent* (*GAC*) if and only if when a variable is assigned any of the values in its domain, there exist compatible values in the domains of all the