Artificial Intelligence 223 (2015) 82–119

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

The deterministic part of the seventh International Planning Competition ☆

Carlos Linares López *, Sergio Jiménez Celorrio, Ángel García Olaya

Computer Science Department, Universidad Carlos III de Madrid, Leganés (Madrid), Spain

ARTICLE INFO

Article history:
Received 22 November 2013
Received in revised form 12 December 2014
Accepted 19 January 2015
Available online 3 February 2015

Keywords:
Automated planning
Planning systems
International Planning Competition
Benchmarks for planning
Experimental evaluation of planning systems

ABSTRACT

The International Planning Competition is organized in the context of the International Conference on Automated Planning and Scheduling (ICAPS) and it is considered a reference source for the planning and scheduling community. The competition is typically organized every two years and deals with relevant issues for the community such as the definition of evaluation standards, the publication of benchmarks and the collection and dissemination of data about state-of-the-art planners. This paper focuses on the deterministic part, the longest-running part of the International Planning Competition. The paper describes its format, the participants, the selection of benchmarks and the generated results accompanied with analysis from different perspectives. The paper also examines the results of a brand new track created to explore the potential of planners that exploit the power of multi-core processors. Overall, the results of the competition indicate significant progress with respect to previous competitions, but they also reveal that some issues remain open and need further research, such as the coverage of temporal planners when concurrency is required and the performance in the multi-core track. As a novelty, all the data and the software generated for running the competition have been made publicly available allowing researchers to reproduce the competition and to carry out different analysis of the results.

1. Introduction

The International Planning Competition (IPC) has typically been held every 2 years in the context of the International Conference on Automated Planning and Scheduling (ICAPS). The main activity in the competition is evaluation of state-of-the-art planners by running them on a set of planning problems and comparing their performance with some specific metrics. Although the IPC is competitive and awards planners for their performance, the main goals of the competition are collection and dissemination of data and definition of evaluation methodologies. In fact, the IPC is used as a reference source when building a planner, and most new planning techniques are evaluated by considering languages, benchmarks, and metrics defined in the IPC series.

Currently the IPC comprises three parts: (1) a deterministic part for evaluation of domain-independent planners in deterministic and fully observable environments; (2) a learning part for planners able to learn and exploit domain-specific knowledge in deterministic planning; and (3) an uncertainty part for domain-independent planners able to plan under un-

☆ This paper was submitted to the Competition Section of the journal.

* Corresponding author.
E-mail addresses: carlos.linares@uc3m.es (C. Linares López), sjimenez@inf.uc3m.es (S. Jiménez Celorrio), agolaya@inf.uc3m.es (Á. García Olaya).

http://dx.doi.org/10.1016/j.artint.2015.01.004
0004-3702/

certainty. The deterministic part has been the main focus of the IPC over the years and is the part addressed in this paper. Details on the other two parts can be found in a survey paper [1] and on the IPC website.[1]

After a 3-year gap, the seventh edition of the deterministic part intentionally provided continuity with the previous competition, IPC-2008, for easier comparisons and to build iteratively on previous work that could benefit the community. For example, IPC-2011 was structured using the same tracks as for IPC-2008 and preserved the same input language and evaluation scores. Likewise, the competition reused domains and problems to better quantify progress in the field. However, IPC-2011 also introduced several new planning domains and a large collection of problems over these domains that can serve as a reference for future research. There was special emphasis on the temporal track, and planning problems that require concurrency (i.e., that at least two different actions have to be performed at the same time) were intentionally included. With the aim of exploring new directions for planning research, the seventh edition introduced a multi-core track that evaluates the performance of planners with algorithms able to exploit multi-processor machines.

There were also similarities and differences with regard to the software used to run the competition. Continuing the work achieved in IPC-2008, there were explicit efforts to highlight the transparency and reproducibility of the results of IPC-2011. This involved the creation of a public repository of all the software, benchmarks, and source codes for the participant planners, along with short papers describing them and all the data generated. This public resource allows researchers to validate the competition results using their own means for further analysis from different perspectives.

The seventh edition of the deterministic part attracted a record number of 55 participants grouped in 31 teams from 11 different countries: Australia, Canada, China, France, Germany, India, Israel, Italy, Spain, UK, and USA. This is almost eight times the number of participants in the first competition, IPC-1998, and double the number in the previous one, IPC-2008. The competition results were presented at a special ICAPS session in July 2011. After the competition, a more detailed analysis was performed to gain more insight into the results and possible causes and consequences. In summary, the main contributions of the deterministic part of IPC-2011 are:

- A collection of new domains and their corresponding problem generators.
- A detailed evaluation of the relative performance of the IPC-2011 planners with respect to new and previously used benchmarks.
- A new track for evaluation of multi-core planners.
- A public repository containing all the data produced during the competition and open-source tools for running the experiments and analyzing the results.
- A detailed analysis of the competition results that provides insights into the performance of state-of-the-art planners from different perspectives such as coverage, quality, CPU time, and memory usage.

The overall aim of this paper is to provide accurate answers to the following questions:

- *What is being measured?* The definition of the scoring schema and how it is affected by the particular goal of every track is discussed in Section 2.3. The characterization of every track is described in Section 2.1. The scoring schema adopted is critically analyzed in Section 6 and various alternatives are examined.
- *What are the benchmarks?* Selection of the benchmark suite is a difficult problem that is addressed in Section 3. The difficulty mainly arises from the need to select challenging yet solvable problems and the fact that it might introduce bias, as discussed in Section 3.2 and Appendix C.1.
- *What are the results?* The main results are presented in Section 4 and are compared with the results from IPC-2008 and closely related tracks in Section 5.
- *General questions and trends* are summarized in Section 7.

The remainder of the paper is organized as follows. Section 2 describes the deterministic part of IPC-2011 in terms of its format, the participants, and the evaluation schema. Section 3 reviews the benchmarks used and details the mechanisms followed in the domain and problem selection. Section 4 analyzes the competition results with respect to coverage, quality, raw speed, and memory usage. Section 5 presents a scalability analysis. The performance of the top planners in IPC-2011 is compared with that of the top planners in IPC-2008, and the performance of parallel solvers is compared with their sequential counterparts. The scoring schema used is critically analyzed in Section 6 and a number of alternatives are discussed. Section 7 summarizes with a number of conclusions. A series of appendices are available online as Supplementary material for further reference. Appendix A describes the participants for all the tracks. Appendix B provides additional details about the selected domains. Appendix C presents results for additional experiments performed once the competition was over. Appendix D shows a novel approach used to select problems that were reused from previous IPCs. Appendix E provides additional details on the way in which various statistical tests were conducted.

[1] http://ipc.icaps-conference.org/.